NAVIGATING SCALING: MODELLING AND ANALYSING

P. Abry[(1)], P. Gonçalvès[(2)]

[(1)] SISYPH, CNRS,
Ecole Normale Supérieure
Lyon, France

[(2)] INRIA Rhône-Alpes,
On Leave at IST-ISR, Lisbon

IN COLLABORATIONS WITH :

P. Flandrin, D. Veitch,
P. Chainais, B. Lashermes, N. Hohn,
S. Roux, P. Borgnat,
M.Taqqu, V. Pipiras, R. Riedi, S. Jaffard

Wavelet And Multifractal Analysis, Cargèse, France, July 2004.

Report Documentation Page

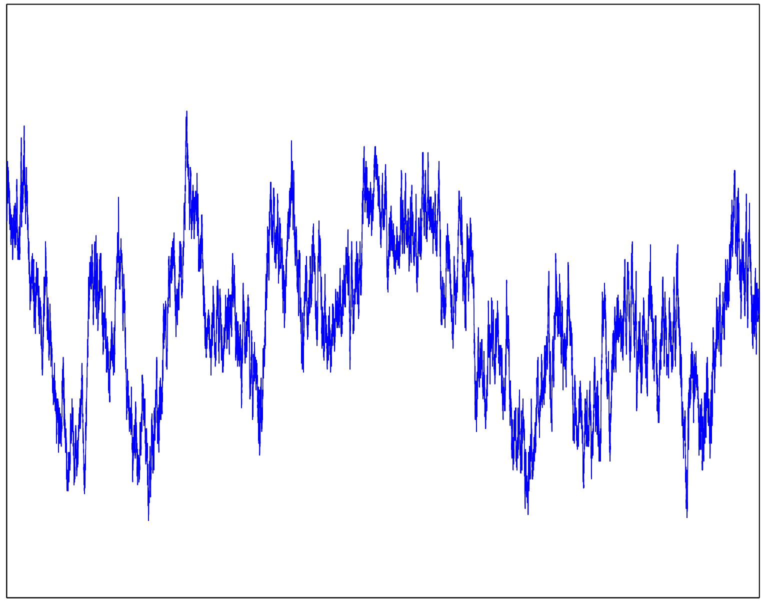
Form Approved
OMB No. 0704-0188

1. REPORT DATE **07 JAN 2005**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Navigating Scaling: Modelling And Analysing**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SISYPH, CNRS, Ecole Normale Sup´erieure Lyon, France**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES **See also ADM001750, Wavelets and Multifractal Analysis (WAMA) Workshop held on 19-31 July 2004., The original document contains color images.**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**64**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

SCALING PHENOMENA ?



- DETECTION: SCALING ? WHAT DOES IT MEAN ? NON STATIONARITY ?
- IDENTIFICATION: RELEVANT STOCHASTIC MODELS ?
- ESTIMATION: RELEVANT PARAMETER ESTIMATION ?
- SIDE ISSUES:

  ROBUSTNESS ? COMPUTATIONAL COST ? REAL TIME ? ON LINE ?

OUTLINE

I. INTUITIONS, MODELS, TOOLS

- I.1 INTUITIONS, DEFINITION,APPLICATIONS
- I.2 STOCHASTIC MODELS: SELF-SIMILARITY VS MULTIFRACTAL
- I.3 MULTIRESOLUTION TOOLS, AGGREGATION, INCREMENTS
- I.4 WAVELETS, CONTINUOUS, DISCRETE

II. SECOND ORDER ANALYSIS, SELF SIMILARITY AND LONG MEMORY

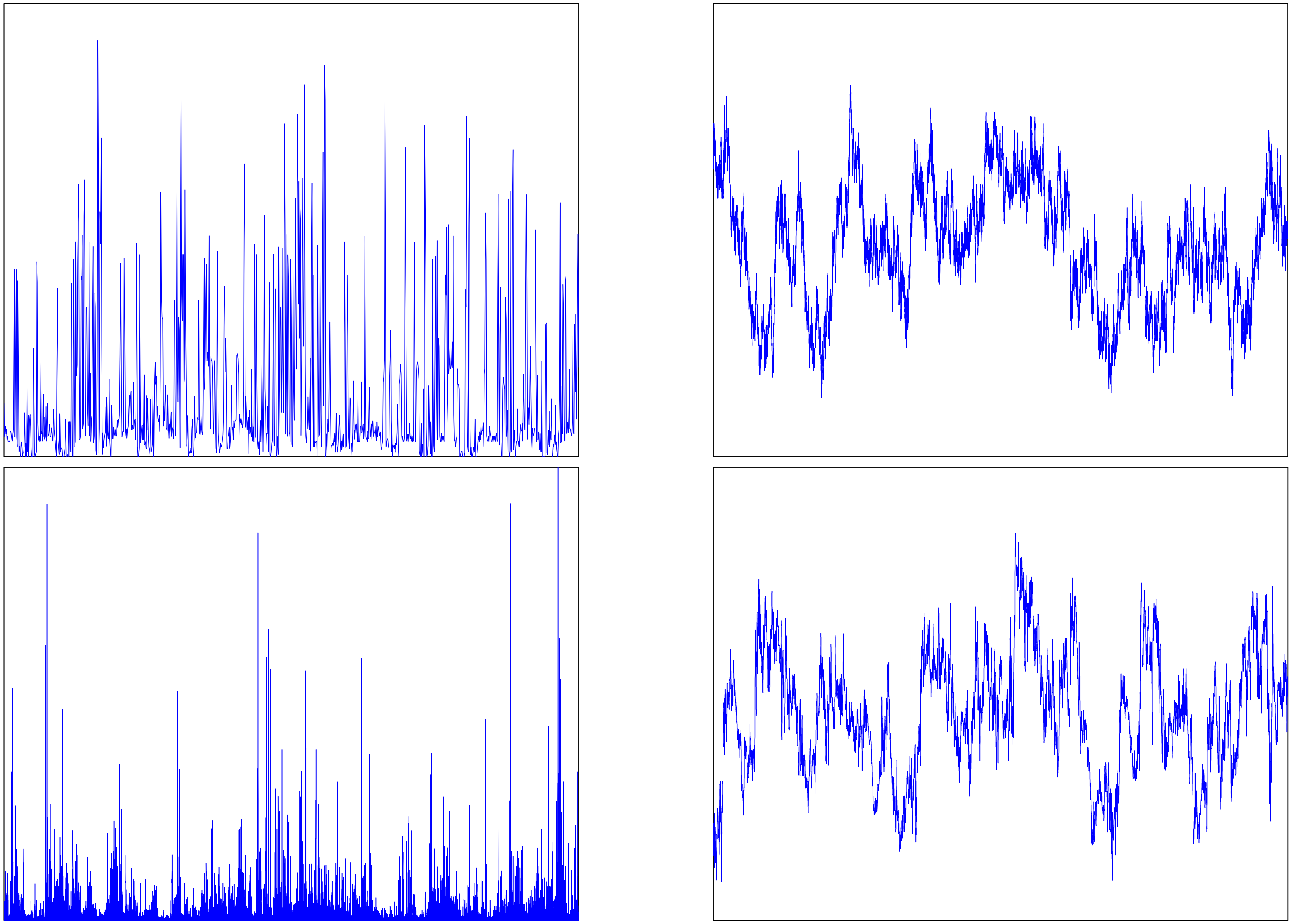
- II.1 RANDOM WAKS, SELF SIMILARITY, LONG MEMORY,
- II.2 2ND ORDER WAVELET STATISTICAL ANALYSIS,
- II.3 ESTIMATION, ESTIMATION PERFORMANCE,
- II.4 ROBUSTNESS AGAINST NON STATIONARITIES,

III. HIGHER ORDER ANALYSIS, MULTIFRACTAL PROCESSES

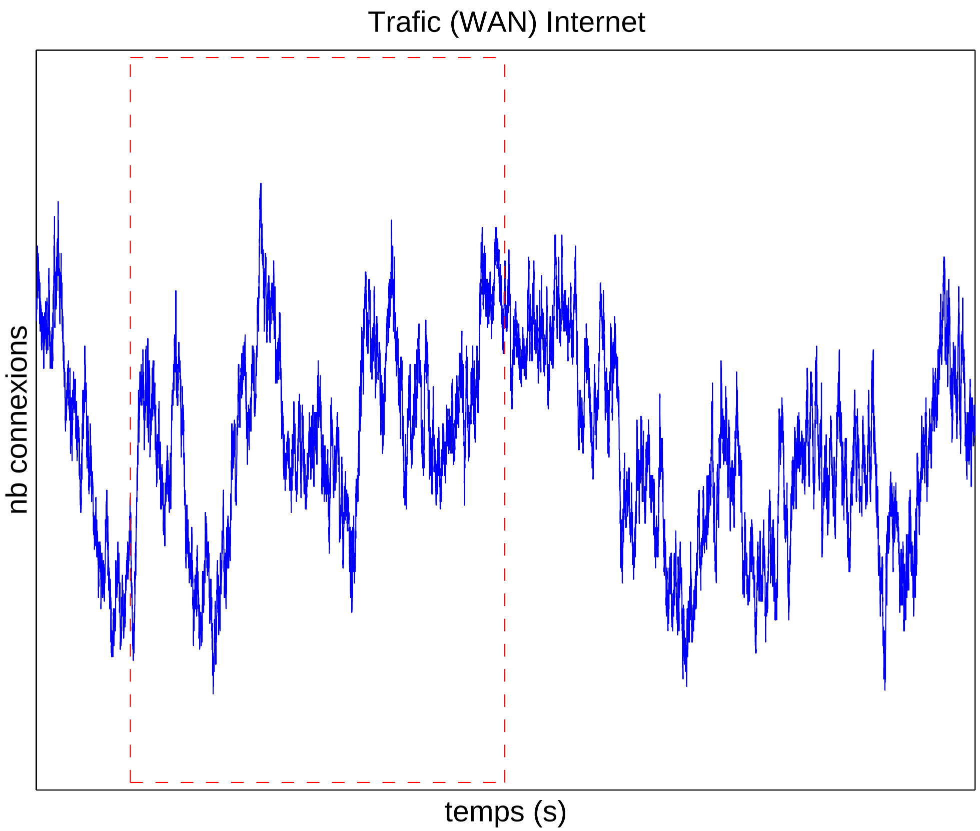
- III.1 MULTIPLICATIVE CASCADES, MULTIFRACTAL PROCESSES,
- III.2 HIGHER ORDER WAVELET STATISTICAL ANALYSIS,
- III.3 FINITENESS OF MOMENTS,
- III.4 ESTIMATION, ESTIMATION PERFORMANCE,
- III.5 NEGATIVE ORDERS,
- III.6 BEYOND POWER LAWS.

IRREGULARITIES, VARIABILITIES

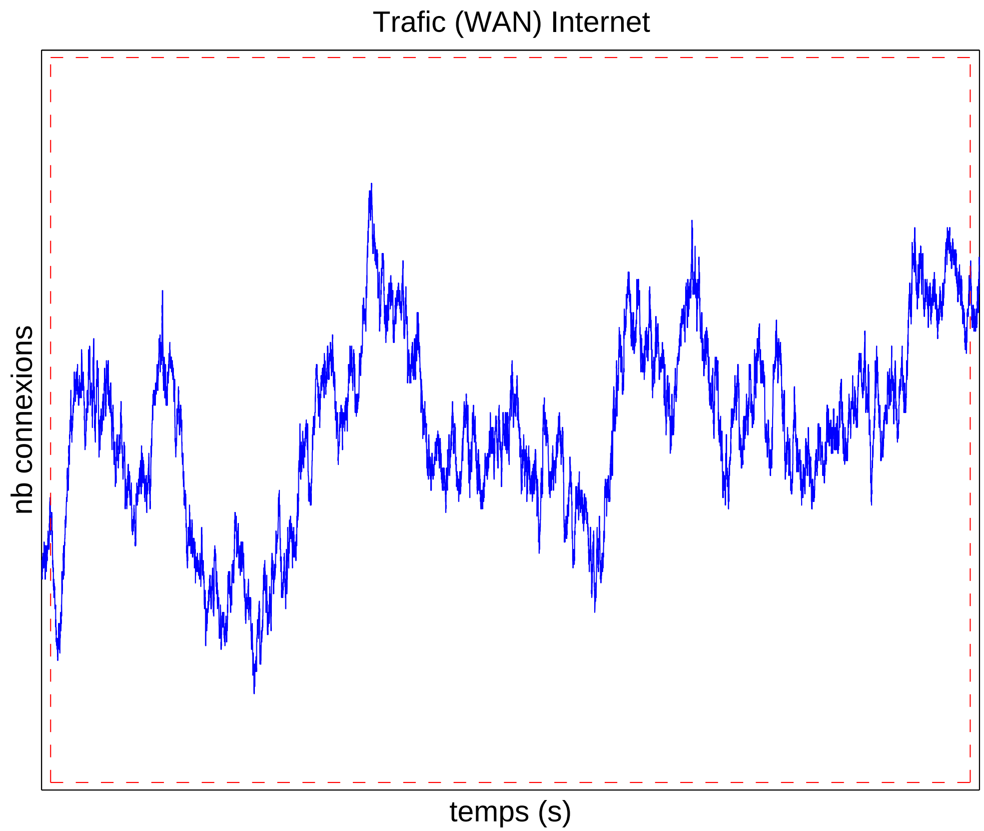
SCALING OR NON STATIONARITIES?



SCALING ?

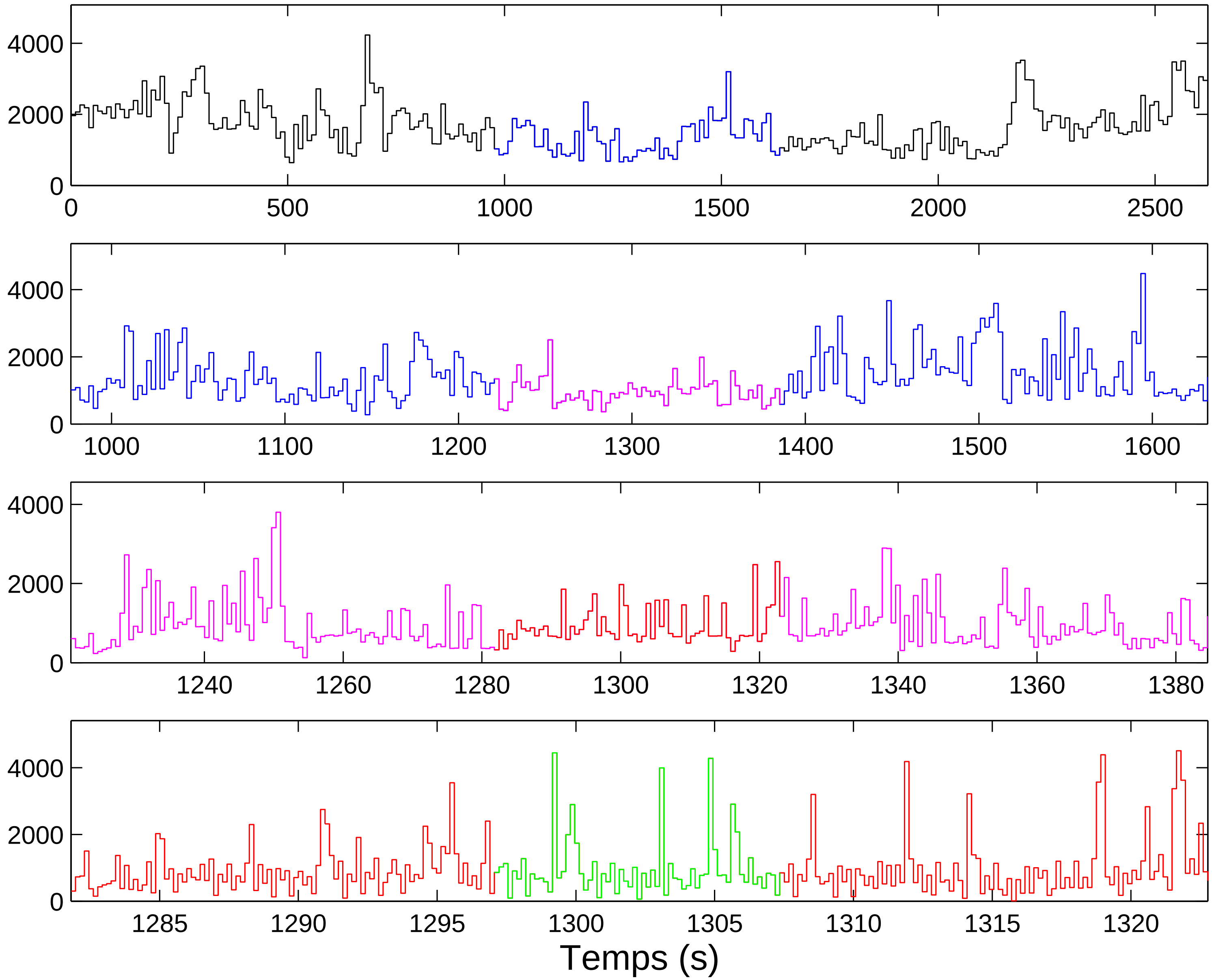


Trafic (WAN) Internet
nb connexions
temps (s)

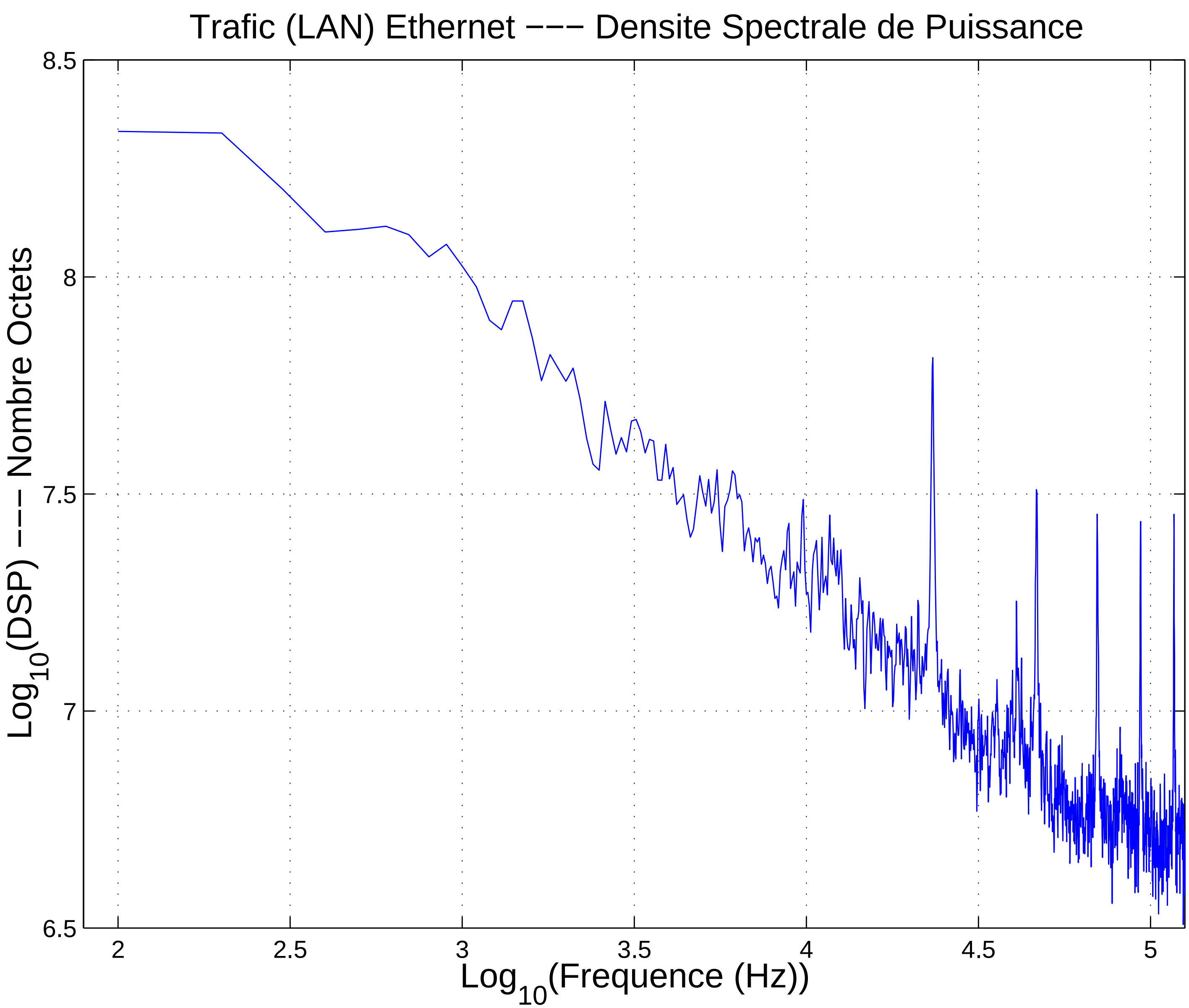


Trafic (WAN) Internet
nb connexions
temps (s)

SCALING ?


4000
2000
0
0
500
1000
1500
2000
2500
4000
2000
0
1000
1100
1200
1300
1400
1500
1600
4000
2000
0
1240
1260
1280
1300
1320
1340
1360
1380
4000
2000
0
1285
1290
1295
1300
1305
1310
1315
1320
Temps (s)

SCALING ?


Trafic (LAN) Ethernet --- Densite Spectrale de Puissance
8.5
8
7.5
7
6.5
2
2.5
3
3.5
4
4.5
5
Log10(DSP) --- Nombre Octets
Log10(Frequence (Hz))

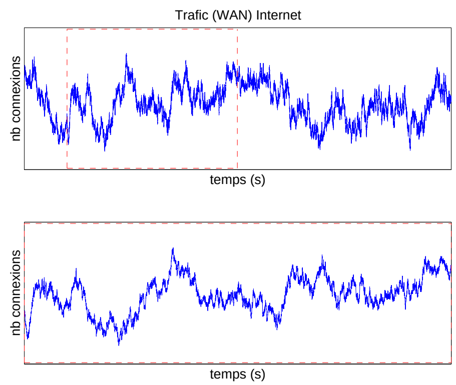
SCALING !

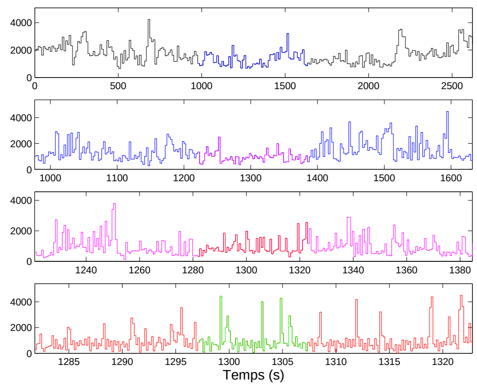
- **DEFINITION :**

NON PROPERTY: NO CHARACTERISTIC SCALE.
NON GAUSSIAN, NON STATIONARY, NON LINEAR

- **EVIDENCE:**

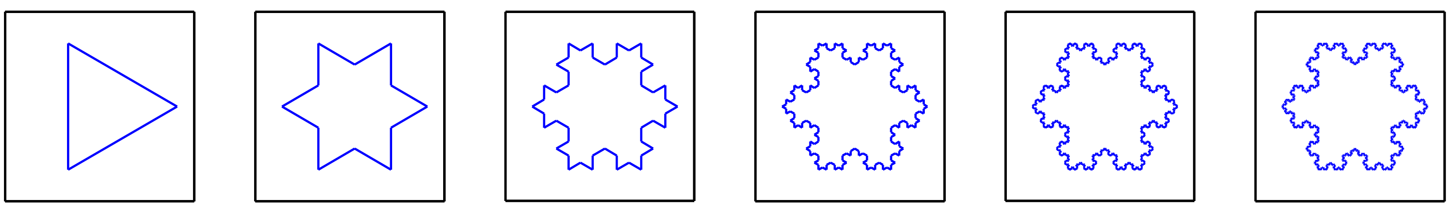
The whole resembles to its part, the part resembles to the whole.





- **ANALYSIS:** Rather than for a characteristic scale,

look for a relation, a mecanism, a cascade between scales.



Scaling : Operational Definitions

- Multiresolution quantity:

$$T_X(a,t) \qquad \text{(e.g., Wavelet Coef.).}$$

- Power Laws:

$$\mathbf{E}|T_X(a,t)|^q = c_q|a|^{\zeta(q)},$$
$$\frac{1}{n}\sum_{k=1}^{n}|T_X(a,t_k)|^q = c_q|a|^{\zeta(q)},$$

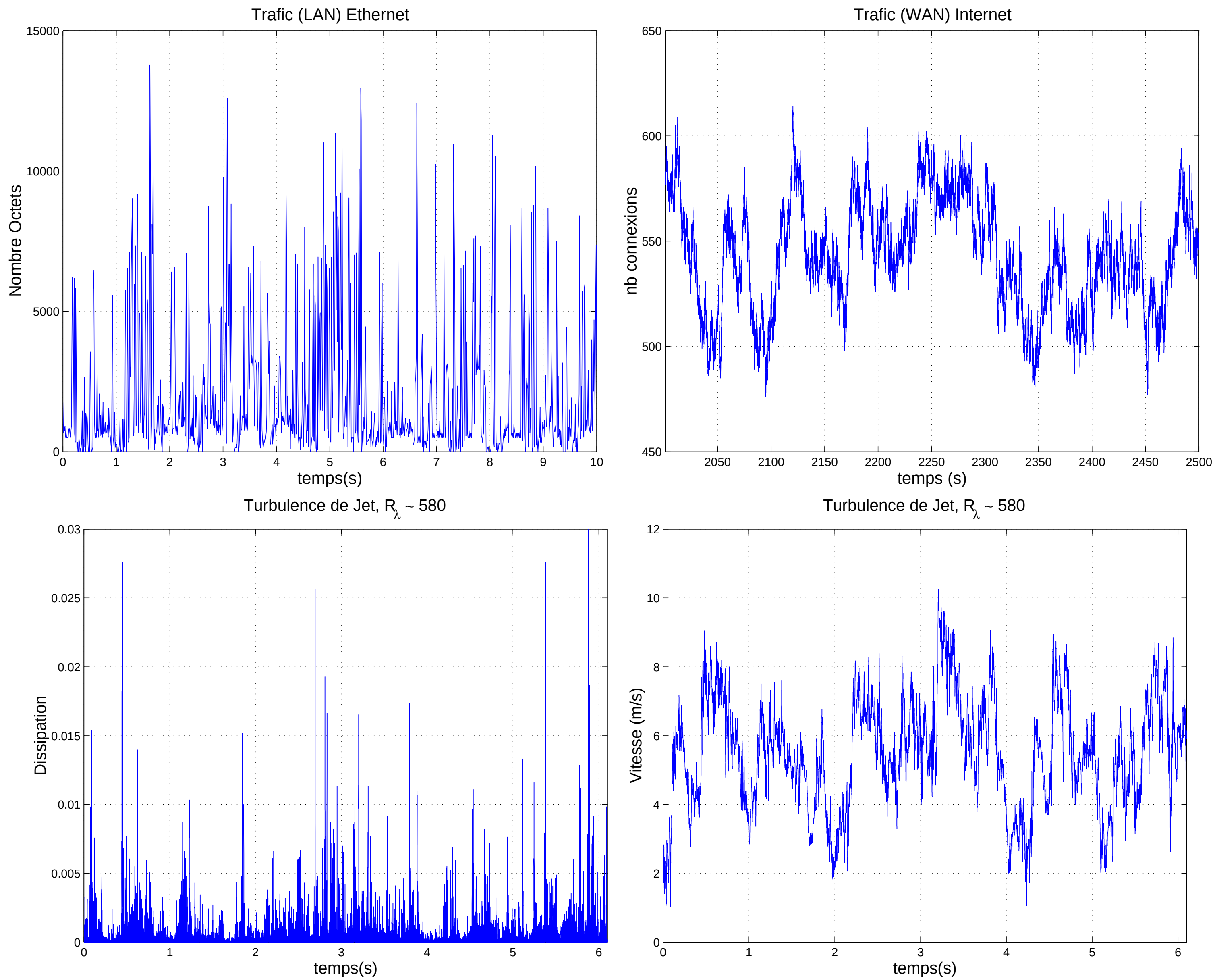
- for a range of scales a,
- for a range of orders q
- Scaling Exponents $\zeta(q)$.

- Beyond Power Laws : Warped Inf. Div. Cascades

$$\mathbf{E}|T_X(a,t)|^q = C_q|a|^{\zeta(q)} = C_q \exp(\zeta(q)\ln a)$$
$$\mathbf{E}|T_X(a,t)|^q = \qquad = C_q \exp(\zeta(q)n(a))$$

$\longrightarrow$ Visit Pierre Chainais's Poster

Ubiquity ?


Trafic (LAN) Ethernet
Nombre Octets
temps(s)
Trafic (WAN) Internet
nb connexions
temps (s)
Turbulence de Jet, Rλ ~ 580
Dissipation
temps(s)
Turbulence de Jet, Rλ ~ 580
Vitesse (m/s)
temps(s)

UBIQUITY !

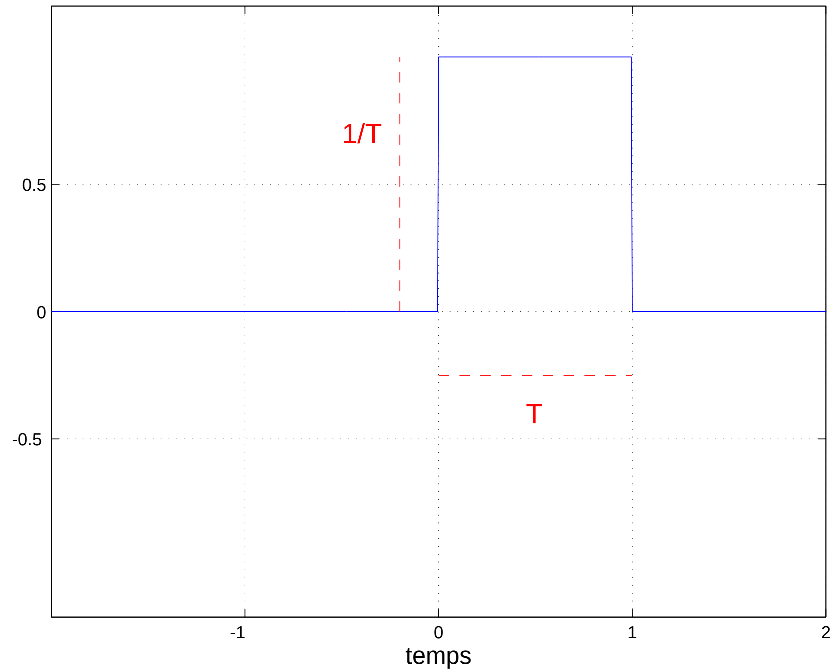
- Hydrodynamic Turbulence,
- Physiology, Biological Rythms (Heart beat, walk),
- Geophysics (Faults Repartition, Earthquakes),
- Hydrology (Water Levels),
- Statistical Physics (Long Range Interactions),
- Thermal Noises (semi-conductors),
- Information Flux on Networks, Computer Network Traffic,
- Population Repartition (local: cities, global: continent),
- Financial Markets (Daily returns, Volatily, Currencies Exchange Rates),
- . . .

ANALYSING TOOL 1 : AGGREGATION

COMPARE DATA AGAINST A BOX, THEN VARY a

$$T_X(a,t) = \frac{1}{aT_0}\int_t^{t+aT_0} X(u)du$$

AVERAGE



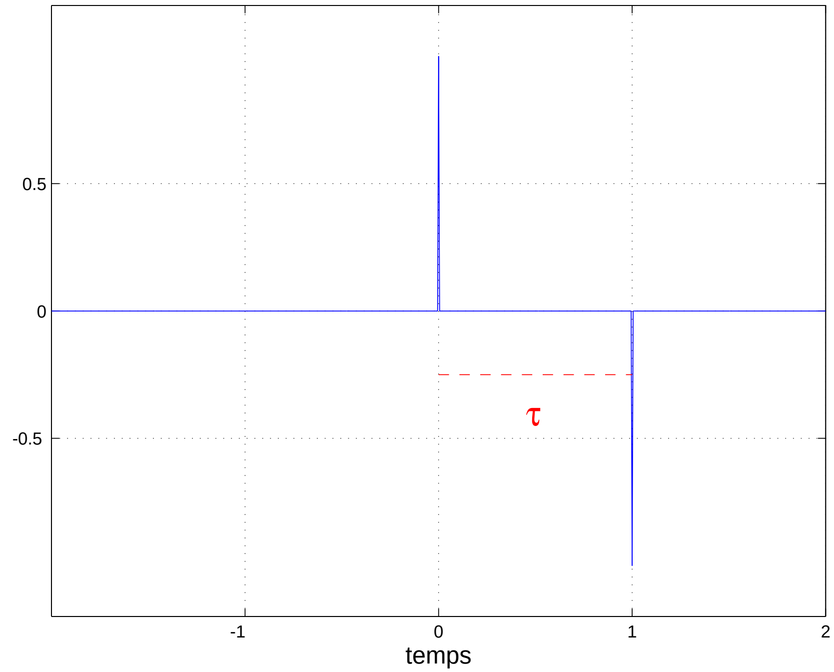
WORKS ONLY FOR POSITIVE TIME SERIES, DENSITY

Analysing Tool 2 : Increments

Compare data against a difference of Delta functions, then vary a

$$T_X(a,t) = X(t + a\tau_0) - X(t)$$

Difference



Increments of higher orders or generalised N-variations

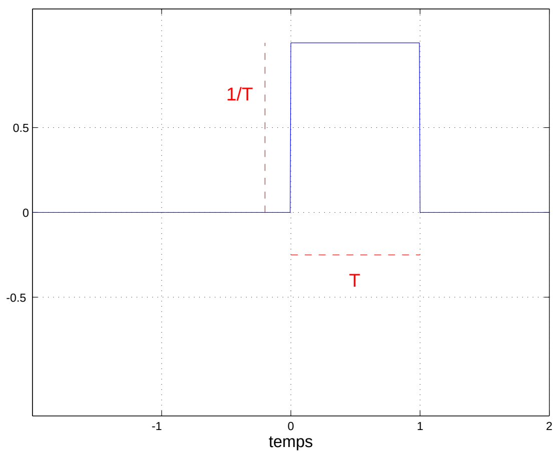
– Order 2 : $T_X(a,t) = -X(t+2a\tau_0) + 2X(t+a\tau_0) - X(t)$,
– Order N : $T_X(a,t) = \sum_{p=0}^{N} (-1)^p a_p X(t + pa\tau_0)$,
where $\sum_{p=0}^{N} (-1)^p a_p p^k \equiv 0, \quad k = 0, \ldots, N-1.$

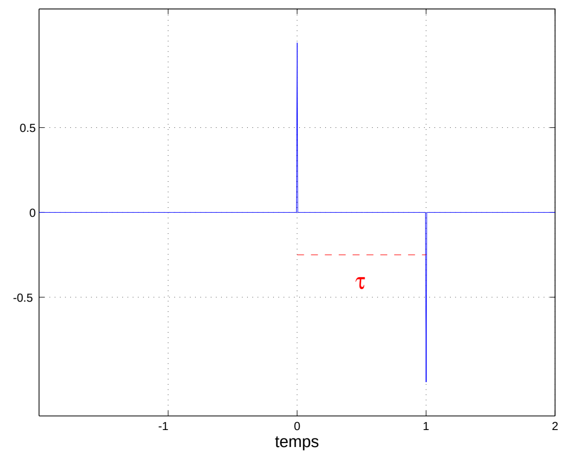
ANALYSING TOOL: MULTIRESOLUTION ANALYSIS

- MULTIRESOLUTION QUANTITIES:

$$X(t) \rightarrow T_X(a,t) = \langle f_{a,t} | X \rangle, \quad f_{a,t}(u) = \frac{1}{a} f_0\left(\frac{u-t}{a}\right)$$

AGGREGATION	INCREMENTS	?
$f_0(u) = (\beta_0)$	$f_0(u) = (I_0)$	?
$= \frac{1}{aT_0} \int_t^{t+aT_0} X(u)du$	$= X(t + a\tau_0) - X(t)$	?
BOX, AVERAGE	DIFFERENCE	?





?

Analysing Tool: MultiResolution Analysis

- **Multiresolution Quantities:**

$$X(t) \to T_X(a,t) = \langle f_{a,t} | X \rangle, \quad f_{a,t}(u) = \frac{1}{a} f_0\left(\frac{u-t}{a}\right)$$

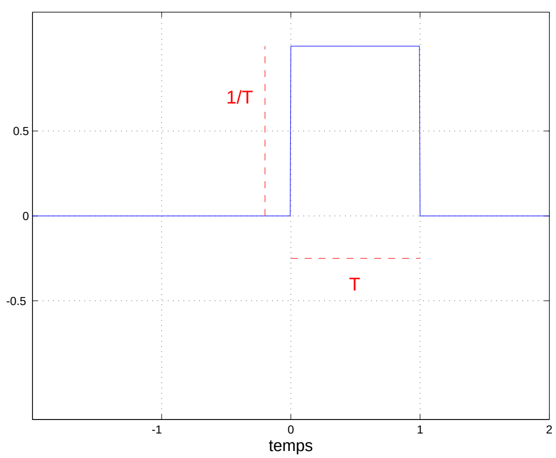
- **Choices for Mother Functions:** f_0,

Aggregation

$$f_0(u) = (\beta_0)^{*N}$$

$$= \frac{1}{aT_0} \int_t^{t+aT_0} X(u)du$$

Box, Average



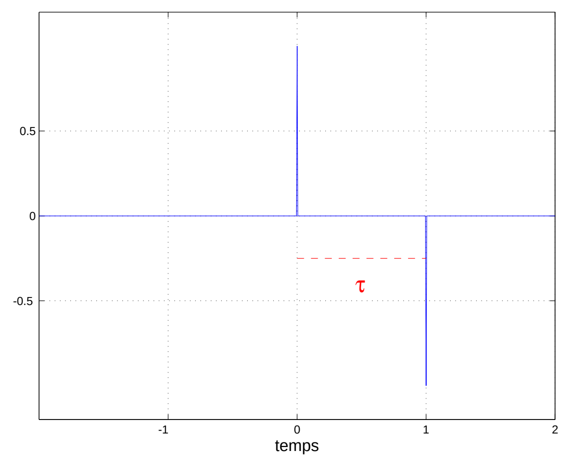
N

Increments

$$f_0(u) = (I_0)^{*N}$$

$$= X(t + a\tau_0) - X(t)$$

Difference



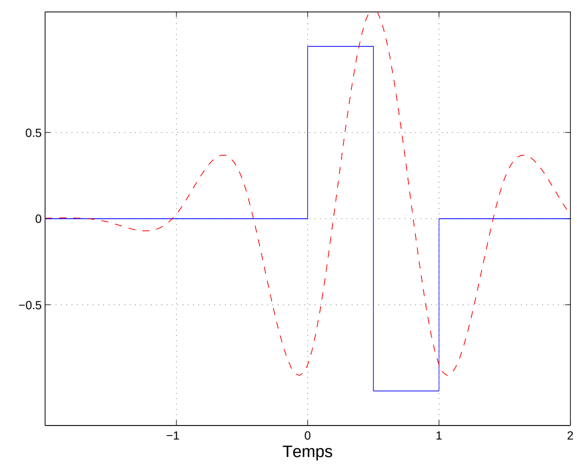
N

Wavelets

$$f_0(u) = \psi_{0,N}$$

$$= \int X(u) \frac{1}{|a|} \psi_0\left(\frac{u-t}{a}\right),$$

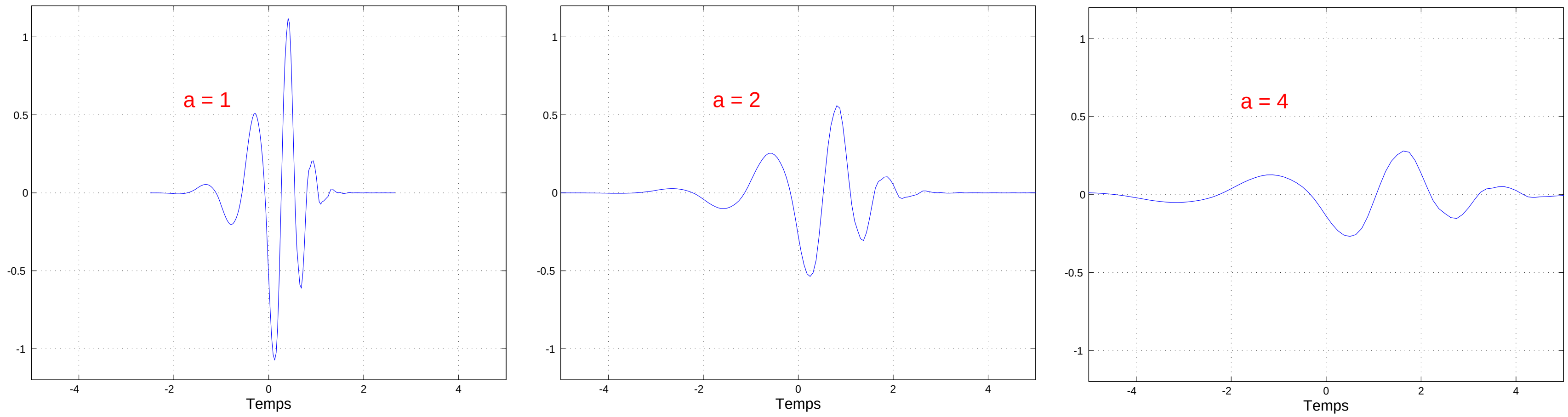
Average, Difference



N

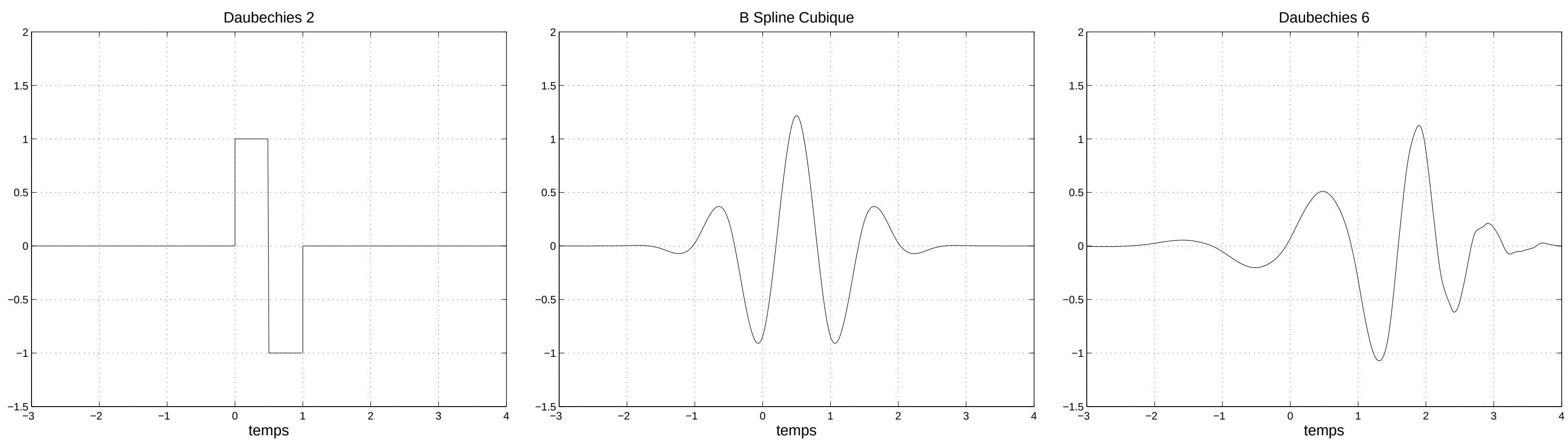
Wavelets and Scaling: Key Ingredients

- **Dilation Operator**, $\frac{1}{|a|}\psi_0(\frac{t}{|a|})$



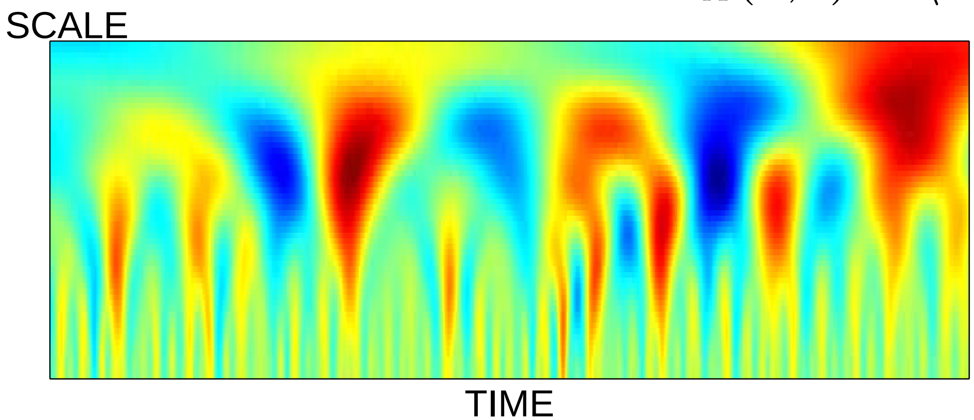
- **Number of Vanishing Moments**,

$$N \geq 1, \int t^k \psi_0(t) dt \equiv 0, \quad k = 0, 1, \ldots, N-1.$$



WAVELET TRANSFORMS

- MOTHER-WAVELET AND "BASIS": $\int \psi_0(u)du = 0, \quad \psi_{a,t}(u) = \frac{1}{|a|}\psi_0(\frac{u-t}{a})$
- WAVELET COEFFICIENTS: CONTINUOUS WT $T_X(a,t) = \langle X, \psi_{a,t} \rangle$

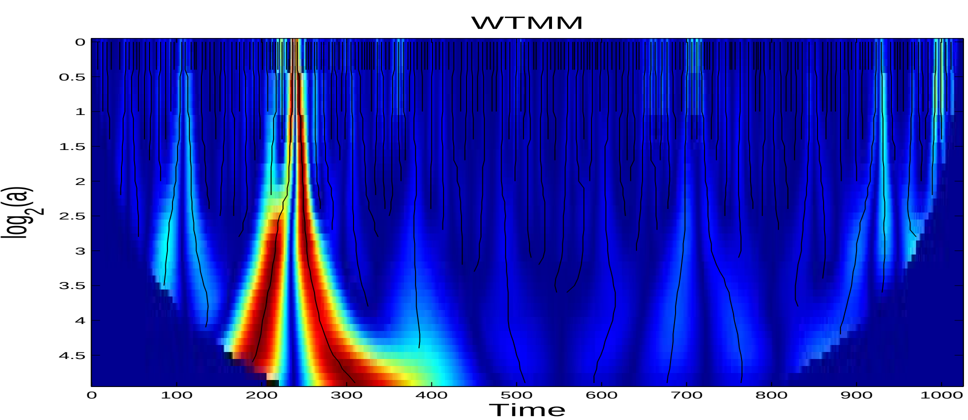


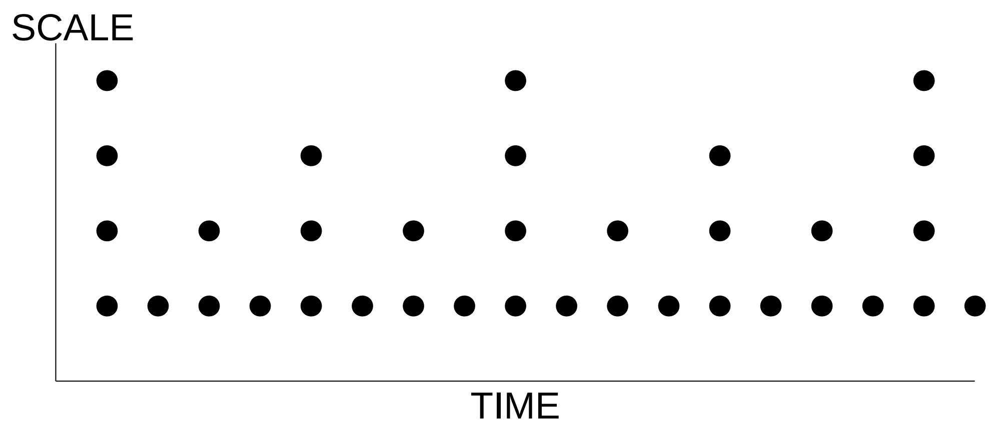
MODULUS MAXIMA WT

SKELETON : MAXIMA LINES

AND DISCRETE WT

$$d_X(j,k) = T_X(a = 2^j, t = 2^j k)$$





Outline

I. Intuitions, models, tools

- I.1 Intuitions, Definition,Applications
- I.2 Stochastic Models: Self-Similarity vs MultiFractal
- I.3 MultiResolution Tools, Aggregation, Increments
- I.4 Wavelets, Continuous, Discrete

II. Second Order Analysis, Self similarity and Long Memory

- II.1 Random Waks, Self Similarity, Long Memory,
- II.2 2nd Order Wavelet Statistical Analysis,
- II.3 Estimation, Estimation Performance,
- II.4 Robustness against Non Stationarities,

III. Higher Order Analysis, Multifractal Processes

- III.1 Multiplicative cascades, Multifractal processes,
- III.2 Higher Order Wavelet Statistical Analysis,
- III.3 Finiteness of Moments,
- III.4 Estimation, Estimation Performance,
- III.5 Negative Orders,
- III.6 Beyond Power Laws.

MOD. TOOL 1: RAND. WALKS AND SELF SIMILARITY

RANDOM WALK: $X(t+\tau) = X(t) + \underbrace{\delta_\tau X(t)}_{\text{Steps or Increments}}$

STATISTICAL PROPERTIES OF THE STEPS:

- **A1:** Stationary,
- **A2:** Independent,
- **A3:** Gaussian,

$\Rightarrow$ Ordinary Random Walk, Ordinary Brownian Motion,

$\Rightarrow$ $\mathbf{E}X(t)^2 = 2D|t|$, Einstein relation,

$\Rightarrow$ $\mathbf{E}|X(t)|^q = 2D|t|^{q/2}$, $q > -1$.

ANOMALIES:

$\Rightarrow$ $\mathbf{E}X(t)^2 = 2D|t|^\gamma$,

$\Rightarrow$ $\mathbf{E}X(t)^2 = \infty$.

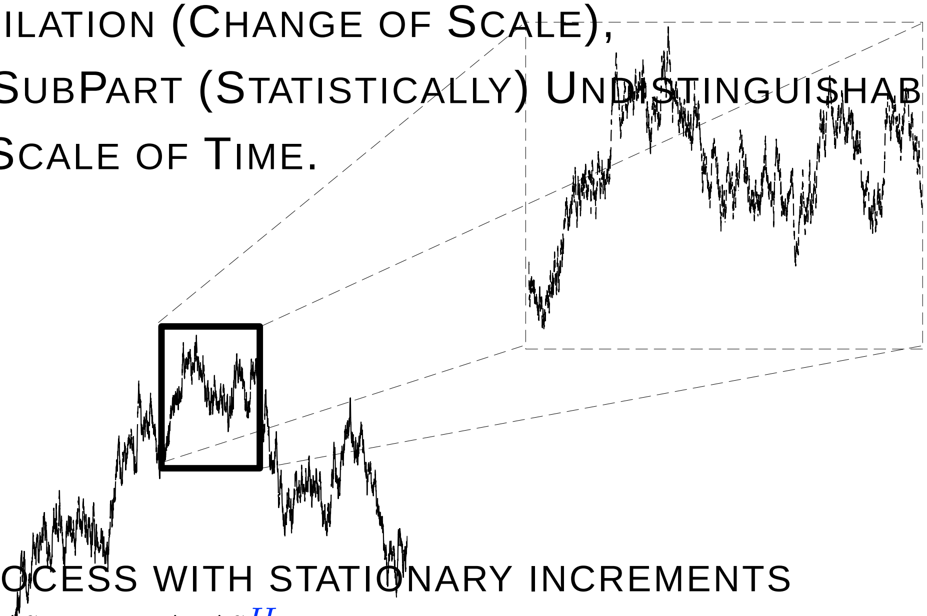
SELF SIMILAR RANDOM WALKS:

- **B1:** Stationary,
- **B2:** Self Similarity

MODELLING TOOL 1: SELF-SIMILARITY

- DEFINITION: $\delta_\tau X(t) \overset{fdd}{=} c^H \delta_{\tau/c} X(t/c), \forall c > 0,$ DILATION FACTOR,
  $1 > H > 0$: SELF-SIMILARITY EXPONENT

- INTERPRETATIONS:
  - COVARIANCE UNDER DILATION (CHANGE OF SCALE),
  - THE WHOLE AND THE SUBPART (STATISTICALLY) UNDISTINGUISHABLE,
  - NO CHARACTERISTIC SCALE OF TIME.



- IMPLICATIONS:
  - NON STATIONARITY PROCESS WITH STATIONARY INCREMENTS
  - $\mathbb{E}|X(t + a\tau_0) - X(t)|^q = C_q |a|^{qH}$,
  - $\forall a > 0, \forall c > 0, \forall q \quad / \mathbb{E}|X(t)|^q < \infty$,
  - A SINGLE SCALING EXPONENT H.
  - ADDITIVE STRUCTURE,
  - (CORRELATED) RANDOM WALK, LONG MEMORY, LONG RANGE CORRELATIONS.

MOD. TOOL 1 (BIS): LONG RANGE DEPENDENCE

- DEFINITIONS :

- LET X BE A 2ND STATIONARY PROCESS WITH,
- COVARIANCE : $c_X(\tau) = \mathbb{E}X(t)X(t+\tau)$
- SPECTRUM : $\Gamma_X(\nu)$

$$\begin{array}{rcl} c_X(\tau) & = & c_\tau |\tau|^{-\beta}, \quad 0 < \beta < 1, \quad |\tau| \to +\infty \\ \Gamma_X(\nu) & = & c_f |\nu|^{-\alpha}, \quad 0 < \alpha < 1, \quad |\nu| \to 0 \end{array}$$

WITH $\alpha = 1 - \beta$ AND $c_f = 2(2\pi)\sin((1-\gamma)\pi/2)c_\tau$.

- CONSEQUENCES :

- $\sum_A^{+\infty} c_X(\tau) d\tau = +\infty,\ A > 0,$
- NO CHARACTERISTIC SCALE,
- AGGREGATION : $T_X(a,t) = \frac{1}{aT_0}\int_t^{t+aT_0} X(u)du,$

$$\Rightarrow \text{VAR}\ T_X(a,t) \sim Ca^{\alpha-1}, \quad a \to +\infty,$$

- INCREMENTS OF SELF.-SIM. PROC. (WITH $H > 1/2$) ARE LONG RANGE DEP. (WITH $\alpha = 2H - 1$).

WAVELETS AND SELF-SIMILAR PROCESSES WITH STATIONARY INCREMENTS - SUMMARY

(Flandrin et al., Tewfik and Kim)

- **P1:** $\{d_X(j,k), k \in \mathcal{Z}\}$ STATIONARY Sequences for each Scale 2^j .
$N \geq 1$

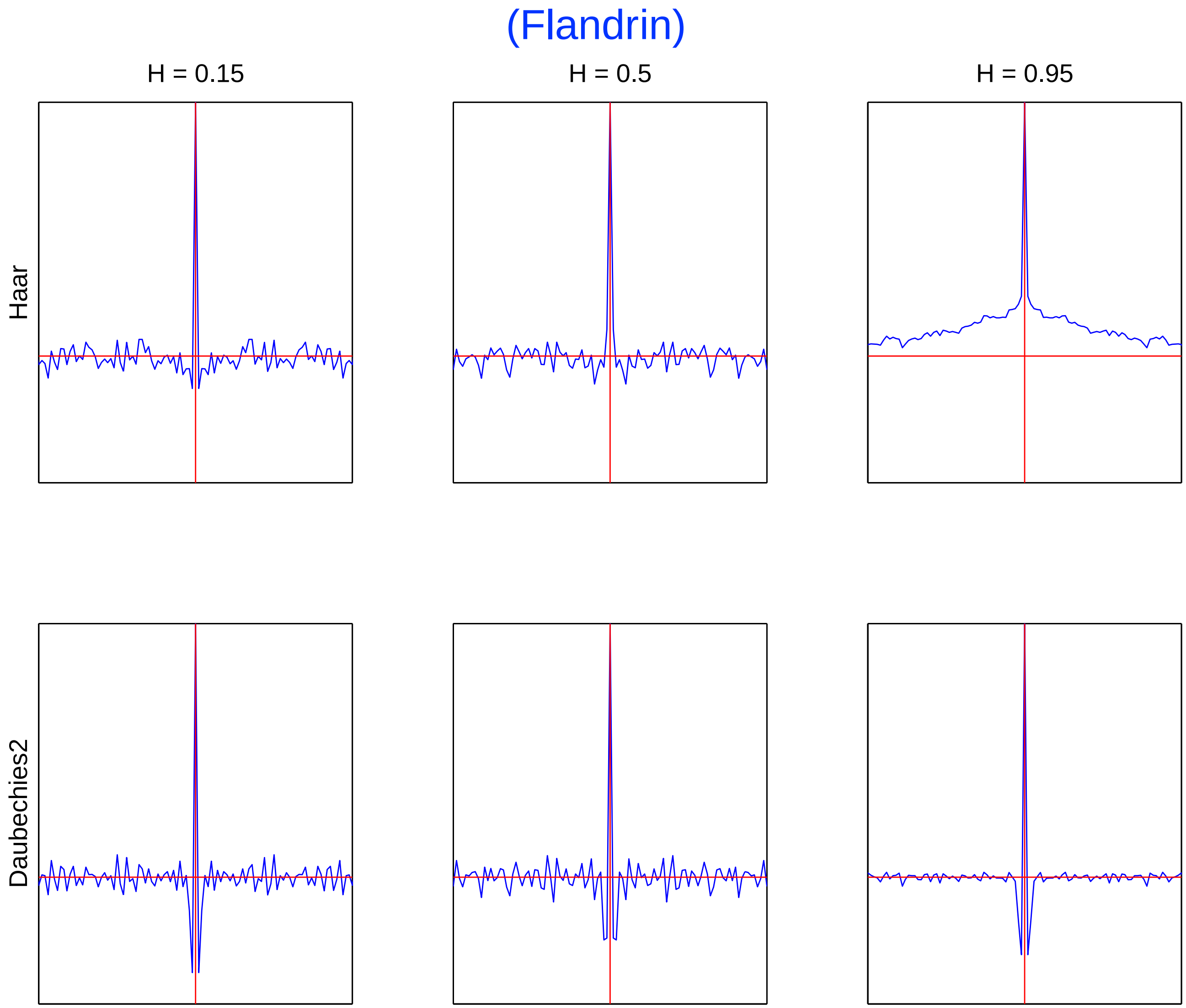
- **P2:** SELF-SIMILARITY : Dilation

$$\{X(t)\} \stackrel{d}{=} \{c^H X(t/c)\} \Rightarrow \{d_X(0,k)\} \stackrel{d}{=} \{2^{-jH} d_X(j,k)\}$$

- P3 : MARGINAL DIST. $P_j(d) = \frac{1}{\beta_0} P_{j'}(\frac{d}{\beta_0}), \quad \beta_0 = \left(\frac{2^{j'}}{2^j}\right)^H$.

- **P4 :** $\{d_X(j,k)\}$ SHORT RANGE DEPENDENT IF $N > H + 1/2$.
$|2^j k - 2^{j'} k'| \to +\infty, \ |\mathrm{Cov}\, d_X(j,k) d_X(j',l)| \leq D |2^j k - 2^{j'} k'|^{2(H-N)}$,
$N \geq 1$ and Dilation

WAVELETS AND LONG RANGE DEPENDENCE



Wavelets and Self-Similar Processes with Stationary Increments - Summary

- **P1:** $\{d_X(j,k), k \in \mathcal{Z}\}$ Stationary Sequences for each Scale 2^j .
$N \geq 1$

- **P2:** Self-Similarity : Dilation

$$\{X(t)\} \stackrel{d}{=} \{c^H X(t/c)\} \Rightarrow \{d_X(0,k)\} \stackrel{d}{=} \{2^{-jH} d_X(j,k)\}$$

- P3 : Marginal Dist. $P_j(d) = \frac{1}{\beta_0} P_{j'}(\frac{d}{\beta_0}), \quad \beta_0 = \left(\frac{2^{j'}}{2^j}\right)^H$.

- **P4 :** $\{d_X(j,k)\}$ Short Range Dependent if $N > H + 1/2$.
$|2^j k - 2^{j'} k'| \to +\infty, \ |\mathrm{Cov}\, d_X(j,k) d_X(j',k')| \leq D |2^j k - 2^{j'} k'|^{2(H-N)}$,
$N \geq 1$ and Dilation

$\Rightarrow$ Idealisation : $d_X(j,k)$ Independent Variables .

$\Rightarrow$ Interpretations: $X(t) = \sum_k a_X(J,k) \varphi_{J,k}(t) + \sum_{j=1,\dots,J,k} d_X(j,k) \psi_{j,k}(t)$.

$\Rightarrow$ Implications: $\mathbf{E}|d_X(j,k)|^q = \mathbf{E}|d_X(0,k)|^q 2^{jqH} \quad \forall q / \mathbf{E}|d_X(0,k)|^q < \infty$.

WAVELETS AND LONG RANGE DEPENDENCE

• SPECTRAL ANALYSIS :
Let X be a 2nd Order stationary process,
Let Ψ be the FT of ψ with central frequency ν_0 and bandwith $\Delta\nu_0$.

$$\begin{aligned} \mathbf{E}|d_X(j,k)|^2 &= \int \Gamma_X(\nu)|\Psi(2^j\nu)|d\nu \\ &\simeq 2^{-j}\Gamma_X(2^{-j}\nu_0) \text{ within bandwith } 2^{-j}\Delta\nu_0. \end{aligned}$$

• LET X BE LONG RANGE DEPENDENT :
- POWER LAW: $\Gamma_X(\nu) = c_f|\nu|^{-\alpha}, 0 < \alpha < 1, |\nu| \to 0$
- POWER LAW: $\mathbf{E}|d_X(j,k)|^2 \sim C2^{j(\alpha-1)}$, $j \to +\infty$,

• $\{d_X(j,k)\}$ SHORT RANGE DEPENDENT IF $N > \alpha - 1$.
$|2^jk - 2^{j'}k'| \to +\infty$, $|\mathrm{Cov}\, d_X(j,k)d_X(j',k')| \leq D|2^jk - 2^{j'}k'|^{\alpha-1-2N}$,
$N \geq 1$ and Dilation

2ND ORDER WAVELET STATISTICAL ANALYSIS

Abry, Gonçalvès, Flandrin

PRINCIPLES:

- IDEAS : **P1** $\Rightarrow \mathbf{E}|d_X(j,k)|^2 = C_2 2^{j2H}$

$\Rightarrow \log_2 \mathbf{E}|d_X(j,k)|^2 = j2H + \beta_q,$

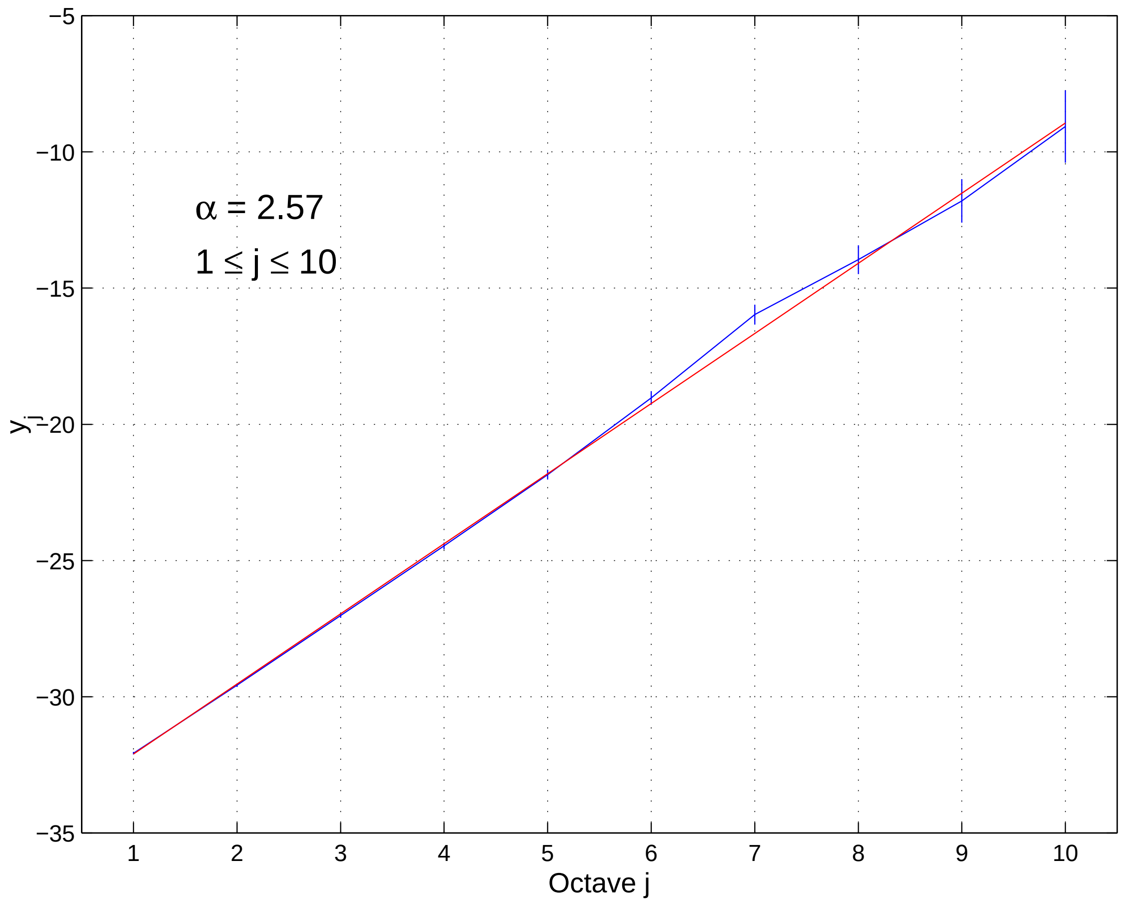
- PROBLEMS: ESTIMATE $\mathbb{E}|d_X(j,k)|^2$ FROM A SINGLE FINITE LENGTH OBSERVATION ?

- SOLUTION : **P2 et P3** $\Rightarrow$ STATISTICAL AVERAGES $\Rightarrow$ TIME AVERAGES,

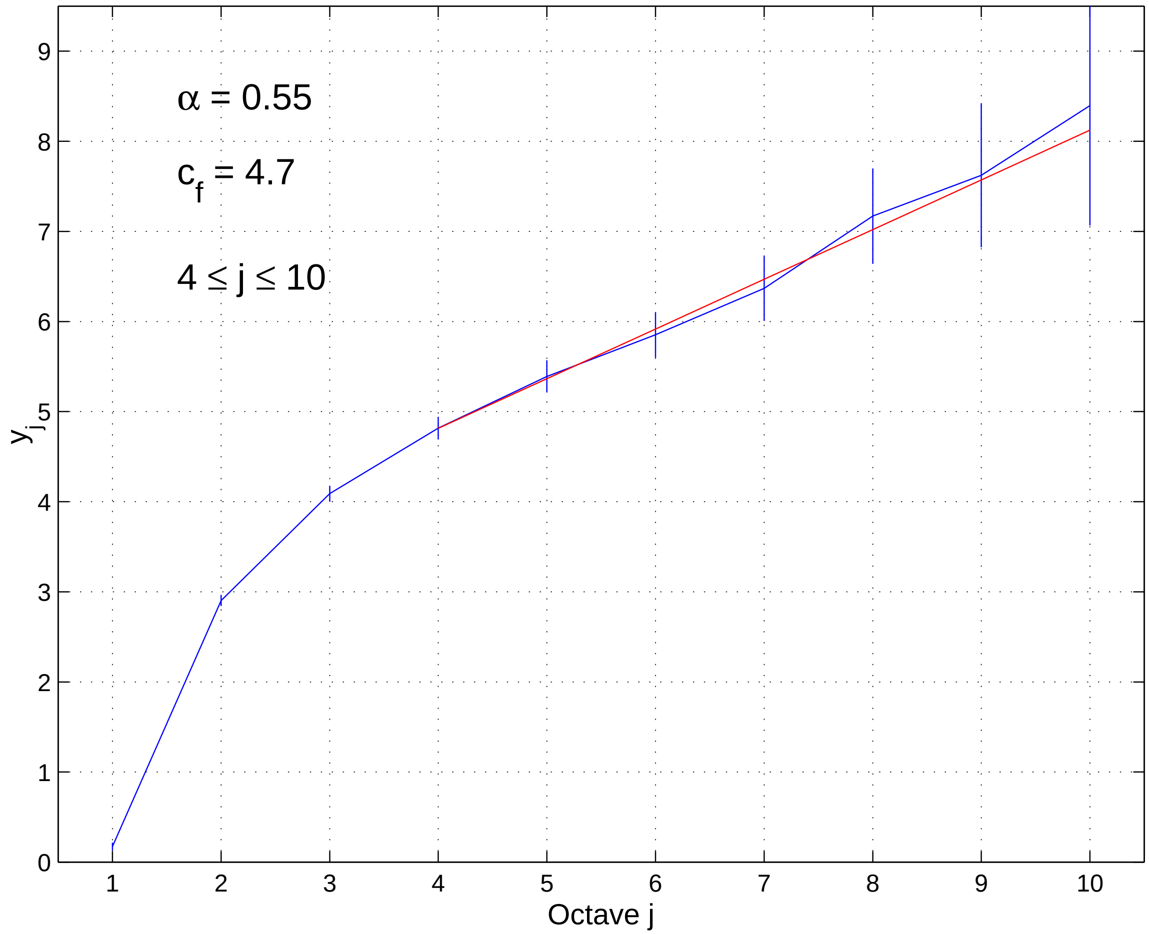
$S_2(j) = (1/n_j)\sum_{k=1}^{n_j} |d_X(j,k)|^2$

LOG-SCALE DIAGRAMS: $\log_2 S_2(j)$ *vs* $\log_2 2^j = j$

2ND ORDER WAVELET-BASED STATISTICAL ANALYSIS FOR SELF-SIMILARITY

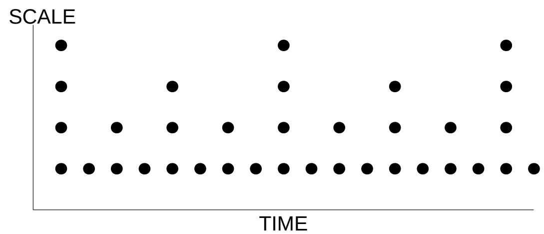


2nd Order Wavelet-based Statistical Analysis for Long Range Dependence



Wavelets and 2nd-order Scaling: Estimation

- Dyadic Grid (Discrete Wavelet Transform): $a_j = 2^j, t_{j,k} = k2^j$,



- Structure Function (Time Average):

$$Y_j = (\tfrac{1}{2}\log_2 S_2(2^j) = \tfrac{1}{2}\log_2(1/n_j)\sum_{k=1}^{n_j}|d_X(j,k)|^2$$

- Definition :

$$Y_j \text{ versus } \log_2 2^j = j,$$
$$\hat{H} = \sum_{j=j_1}^{j_2} w_j Y_j \ .$$

where $\sum_j j w_j \equiv 1, \quad \sum_j w_j \equiv 0$, with $w_j \equiv \frac{B_0 j - B1}{B_0 B_2 - B_1^2}$,
and $\quad p = 0, 1, 2, \quad B_p = \sum_j j^p / a_j, \quad a_j$ arbitrary numbers.

- What Are the Performance of Such an Estimator ?
  When applied to a Self-Similar. or LRD Process

Wavelets and 2nd-order Scaling: Estimation

Abry, Gonçalvès, Flandrin, Abry, Veitch

- Assume: - $i) X$ Gaussian,
  - $ii)$ Idealisation: exact independence.

- Bias : $\mathbf{E}\log_2 S_2(j) = \log_2 \mathbf{E} S_2(j) + \underbrace{\Gamma'(n_j/2) - \log_2(n_j/2)}_{g_j}.$

$$\Rightarrow \mathbf{E}\hat{H} = H + \frac{1}{2}\sum_j w_j g_j.$$

- Variance: $-\text{Var}\,\hat{H} = \frac{1}{4}\sum_j w_j^2 \sigma_j^2,$
  - $-$ min Var $\hat{H} \Longrightarrow a_j \propto \text{Var}\,\log_2 S_2(j)$
  - $-$ Var $\log_2 S_2(j) \simeq C/n_j \simeq 2^j C/n,$
  - $\Rightarrow \text{Var}\,\hat{H} \simeq \left((\log_2 e)^2(\sum_j w_j^2 2^j)\right)/n,$
  - $\Rightarrow$ Analytical (approximate) Confidence Interval (does not depend on unknown H).

- Actual Performances : Negligible Biais, extremely close to MLE.

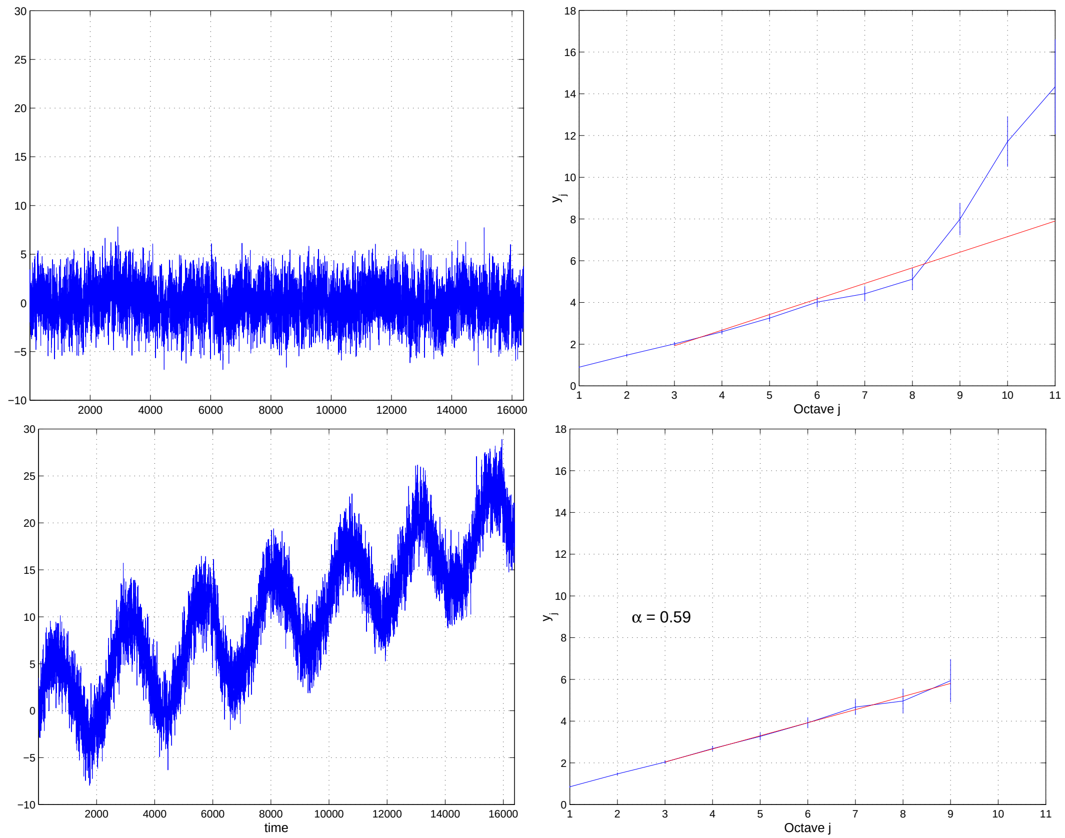
- Conceptual and Practical Simplicity : Matlab Code Available.

Wav. and 2nd-order Scaling: Robustness

Superimposed Trends

$$Y(t) = X(t) + T(t) \Rightarrow d_Y(j,k) = d_X(j,k) + d_T(j,k)$$

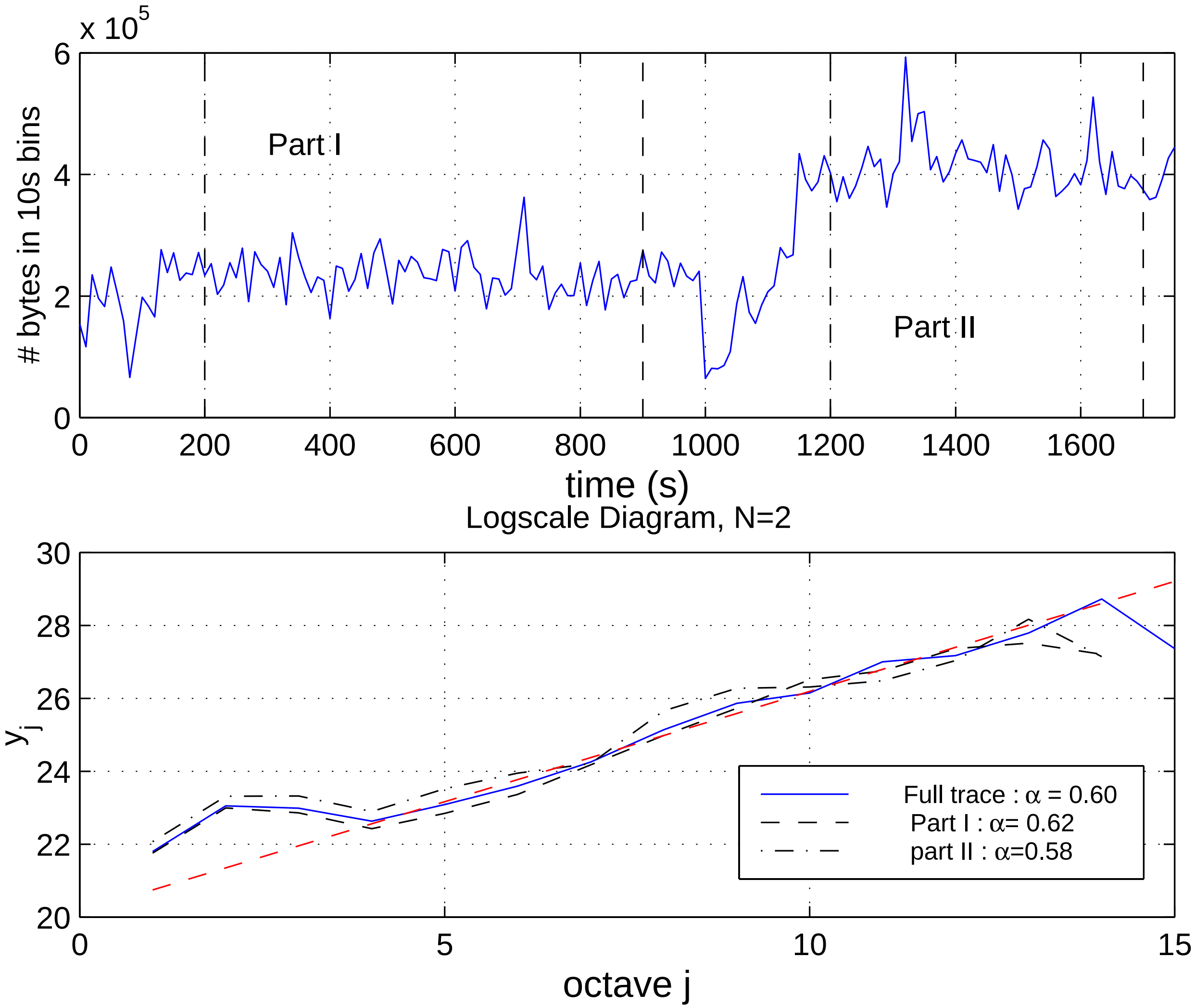
- If $T(t)$ Polynomial of degree P, then $d_T \equiv 0$ when $N > P$,
- If $T(t)$ smooth trend, then the d_T decrease as N increases.



Vary N !

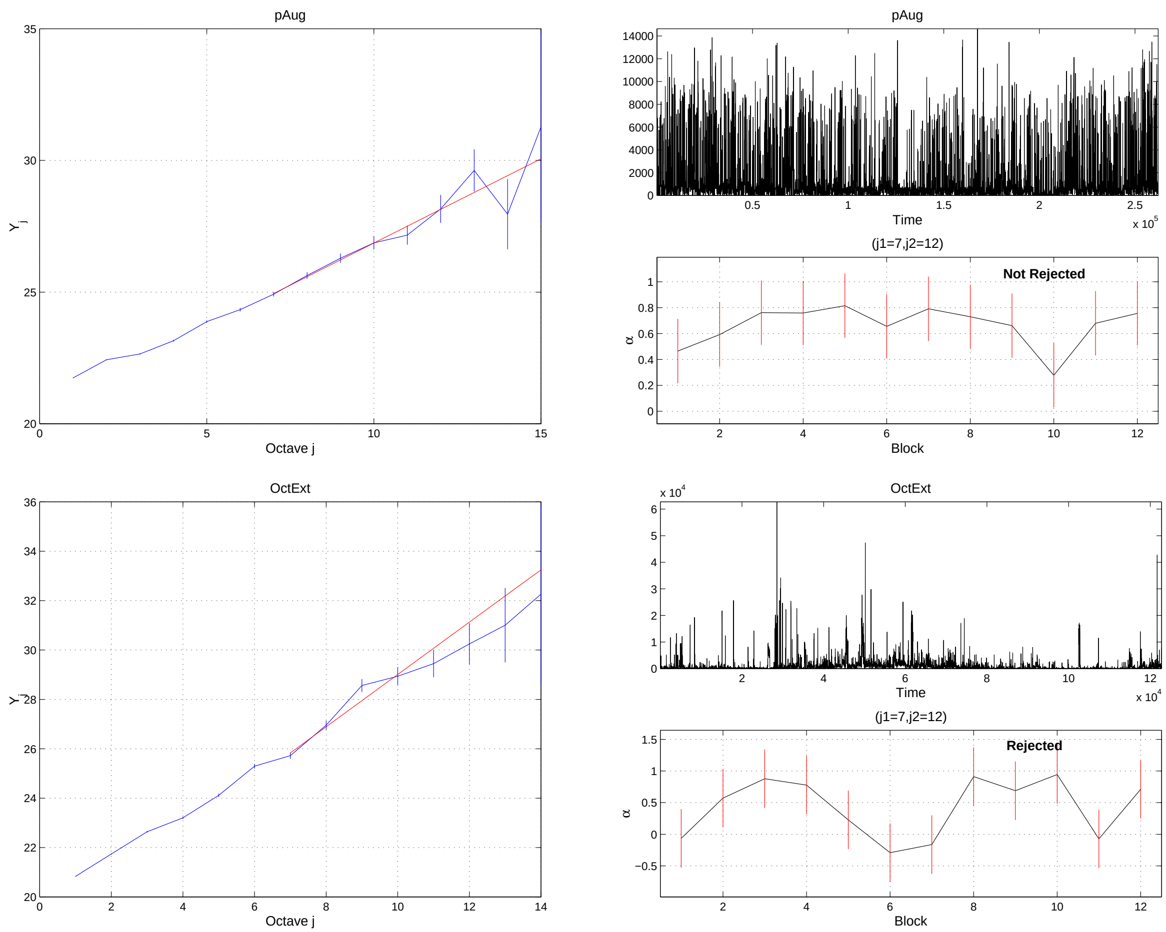
Wav. and 2nd-order Scaling: Robustness

Superimposed Trends - Ethernet Data (Veitch, Abry)



WAV. AND 2ND-ORDER SCALING: ROBUSTNESS

Constancy along time of Scaling laws (Veitch, Abry)



SELF-SIMILARITY

- SELF-SIMILARITY:

$$\mathbf{E}|d_X(j,k)|^q = C_q(2^j)^{qH}$$

  - Power Laws,
  - $\forall 2^j$ (for all scales),
  - $\forall q / \mathbf{E}|d_X(j,k)|^q < \infty$,
  - A single parameter H
  - Additive Structure.

- ?

- ?

Beyond Self-Similarity

- Self-Similarity:

$$\mathbf{E}|d_X(j,k)|^q = C_q(2^j)^{qH}$$

- Power Laws,
- $\forall 2^j$ (for all scales),
- $\forall q / \mathbf{E}|d_X(j,k)|^q < \infty$,
- A single parameter H
- Additive Structure.

- MultiFractal

$$\mathbf{E}|d_X(j,k)|^q = C_q(2^j)^{\zeta(q)}$$

- Power Laws,
- $\forall 2^j < L$, (for fine scales only, in the limit $2^j \to 0$,)
- $\forall q?$
- A whole collection of scaling parameter $\zeta(q)$
- Multiplicative Structure.

- ?

OUTLINE

I. INTUITIONS, MODELS, TOOLS

I.1 INTUITIONS, DEFINITION,APPLICATIONS

I.2 STOCHASTIC MODELS: SELF-SIMILARITY VS MULTIFRACTAL

I.3 MULTIRESOLUTION TOOLS, AGGREGATION, INCREMENTS

I.4 WAVELETS, CONTINUOUS, DISCRETE

II. SECOND ORDER ANALYSIS, SELF SIMILARITY AND LONG MEMORY

II.1 RANDOM WAKS, SELF SIMILARITY, LONG MEMORY,

II.2 2ND ORDER WAVELET STATISTICAL ANALYSIS,

II.3 ESTIMATION, ESTIMATION PERFORMANCE,

II.4 ROBUSTNESS AGAINST NON STATIONARITIES,

III. HIGHER ORDER ANALYSIS, MULTIFRACTAL PROCESSES

III.1 MULTIPLICATIVE CASCADES, MULTIFRACTAL PROCESSES,

III.2 HIGHER ORDER WAVELET STATISTICAL ANALYSIS,

III.3 FINITENESS OF MOMENTS,

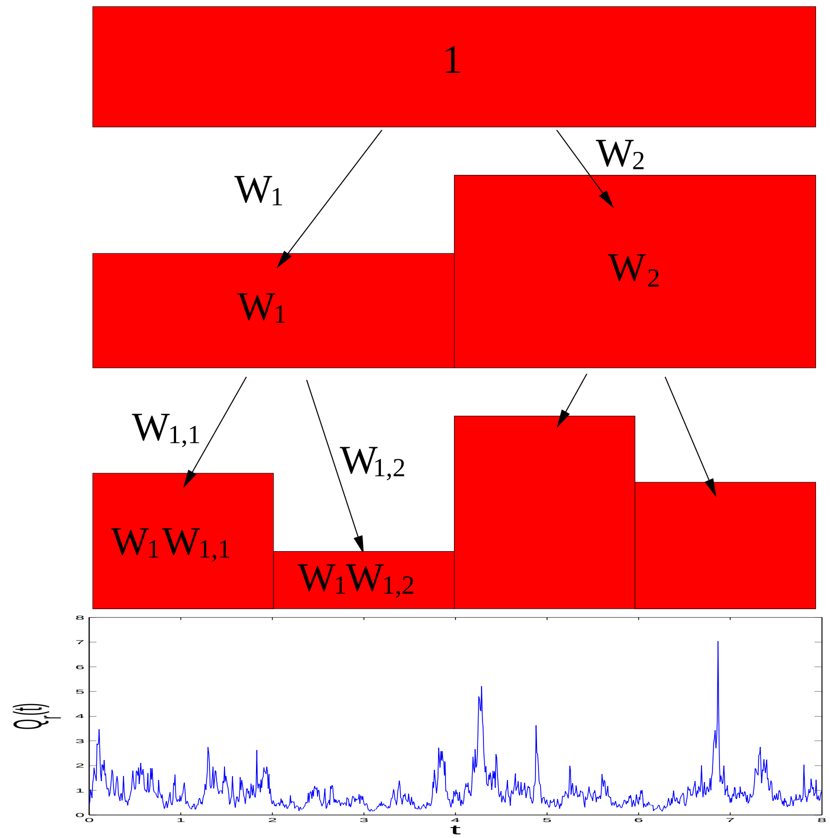
III.4 ESTIMATION, ESTIMATION PERFORMANCE,

III.5 NEGATIVE ORDERS,

III.6 BEYOND POWER LAWS.

Modelling Tool 2: Multiplicative Cascades

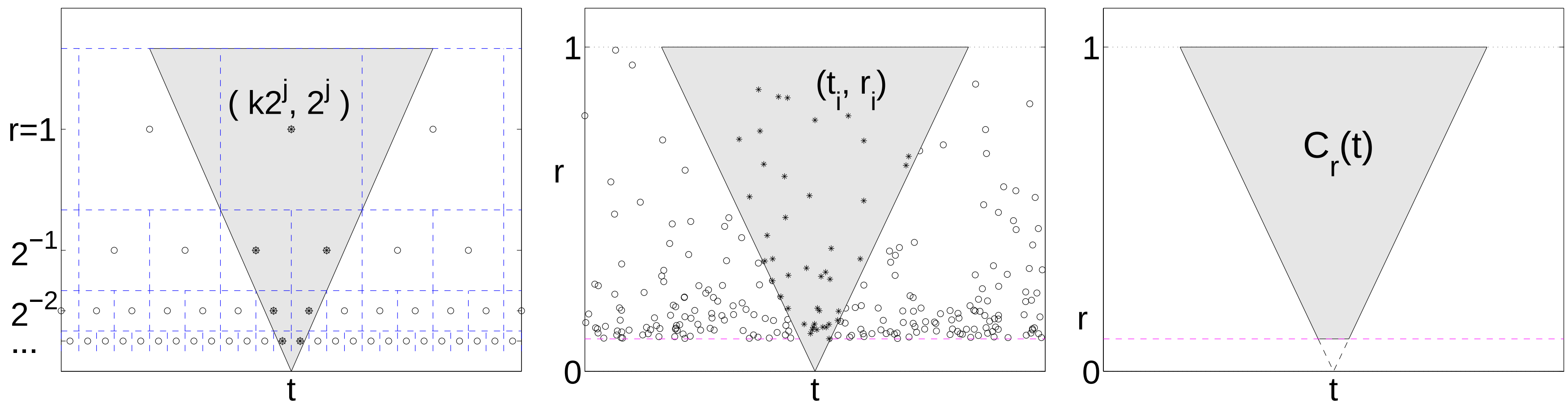
- **Definition:**
  - Split dyadic intervals $I_{j,k}$ into two,
  - I.I.D. multipliers $W_{j,k}$
  - $Q_J(t) = \Pi_{\{(j,k):1 \le j \le J, t \in I_{j,k}\}} W_{j,k}$,



- **Implications:**
  - Local Holder exponent,
  - MultiFractal sample paths, MultiFractal spectrum $D(h)$
  - Cascades, multiplicative structure,
  - $\sum_k \left(1/a \int_{t_k}^{t_k + a\tau_0} X(u)du\right)^q = C_q |a|^{\zeta_q}$, Fine scales $a \to 0$,
  - Multiple exponents ζ_q,
  - No characteristic scale,
  - $\zeta_q = -\log_2 \mathbb{E} W^q$, Non linear in q.

Modelling Tool 2: Multiplicative Cascades

Yaglom, Mandelbrot	Barral, Mandelbrot	Schmitt et al., Bacry et al., Chainais et al.
Mandelbrot's Cascade (CMC)	Compound Poisson Cascade (CPC)	Infinitely Divisible Cascade (IDC)
- IID W,	- IID W,	- Continuous Infinitely
- Dyadic Grid,	- Point Process,	- Divisible Measure M,



$$Q_r(t) = \Pi\ W_{j,k}, \qquad \Pi\ W_{j,k}, \qquad \exp \int dM(t', r'),$$

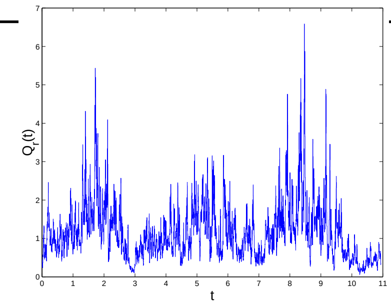
$$\varphi(q) = -\log_2 \mathbb{E} W^q, \qquad = -q(1 - \mathbb{E}W) + 1 - \mathbb{E}W^q, \qquad = \rho(q) - q\rho(1),$$

$$A(t) = \lim_{r \to 0} \int_0^t Q_r(u) du,$$

For a Range of qs, $\mathbb{E}|A(t + a\tau_0) - A(t)|^q = c_q |a|^{q+\varphi(q)}$,

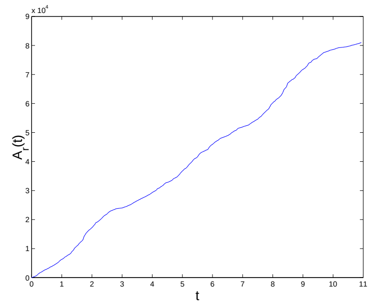
Resolution Depth < Scale < Integral Scale, $a_m = r < a < a_M = L$.

MultiFractal Processes



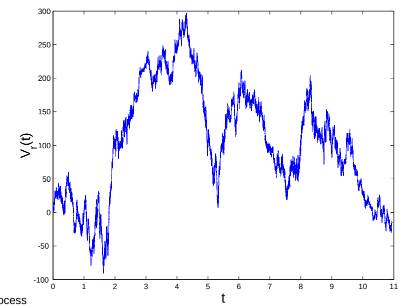
Density: $Q_r(t) = \Pi \; W_{j,k}$

$$\mathbb{E}\left(\frac{1}{a}\int_t^{t+a\tau_0} Q_r(u)du\right)^q = c_q a^{\varphi(q)},$$



Measure: $A(t) = \lim_{r\to 0}\int_0^t Q_r(u)du,$

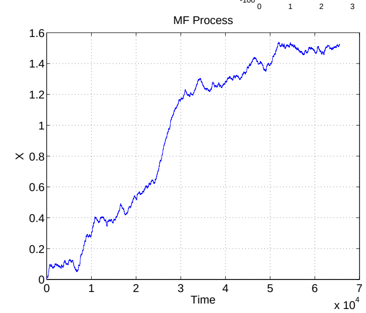
$$\mathbb{E}|A(t+a\tau_0) - A(t)|^q = c_q|a|^{q+\varphi(q)},$$



Fractional Brownian Motion in MultiFractal Time:

$$V_H(t) = B_H(A(t)),$$

$$\mathbb{E}|V_H(t+a\tau_0) - V_H(t)|^q = c_q|a|^{qH+\varphi(qH)},$$



MultiFractal Random Walk:

$$Y_H(t) = \int^t Q_r(s)dB_H(s),$$

$$\mathbb{E}|Y_H(t+a\tau_0) - Y_H(t)|^q = c_q|a|^{qH+\varphi(q)}.$$

Matlab Synthesis Routines : Chainais, Abry

HIGHER-ORDER WAVELET STATISTICAL ANALYSIS

PRINCIPLES :

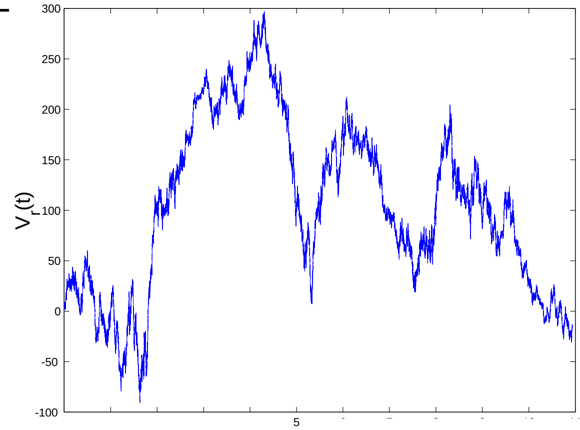
- IDEAS : **P1** $\Rightarrow \mathbb{E}|d_X(j,k)|^q = \mathbb{E}|d_X(0,k)|^q 2^{j\zeta_q}$
$\Rightarrow \log_2 \mathbb{E}|d_X(j,k)|^q = j\zeta_q + \beta_q,$

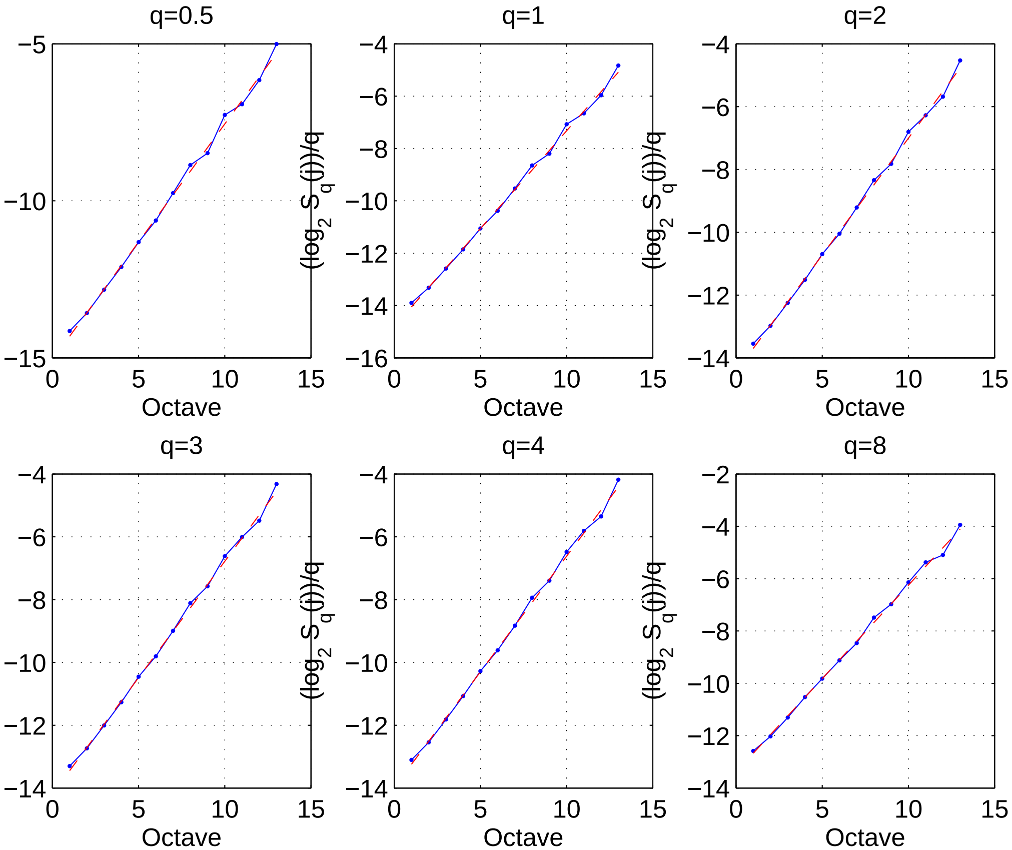
- PROBLEMS: ESTIMATE $\mathbb{E}|d_X(j,k)|^q$ FROM A SINGLE FINITE LENGTH OBSERVATION ?

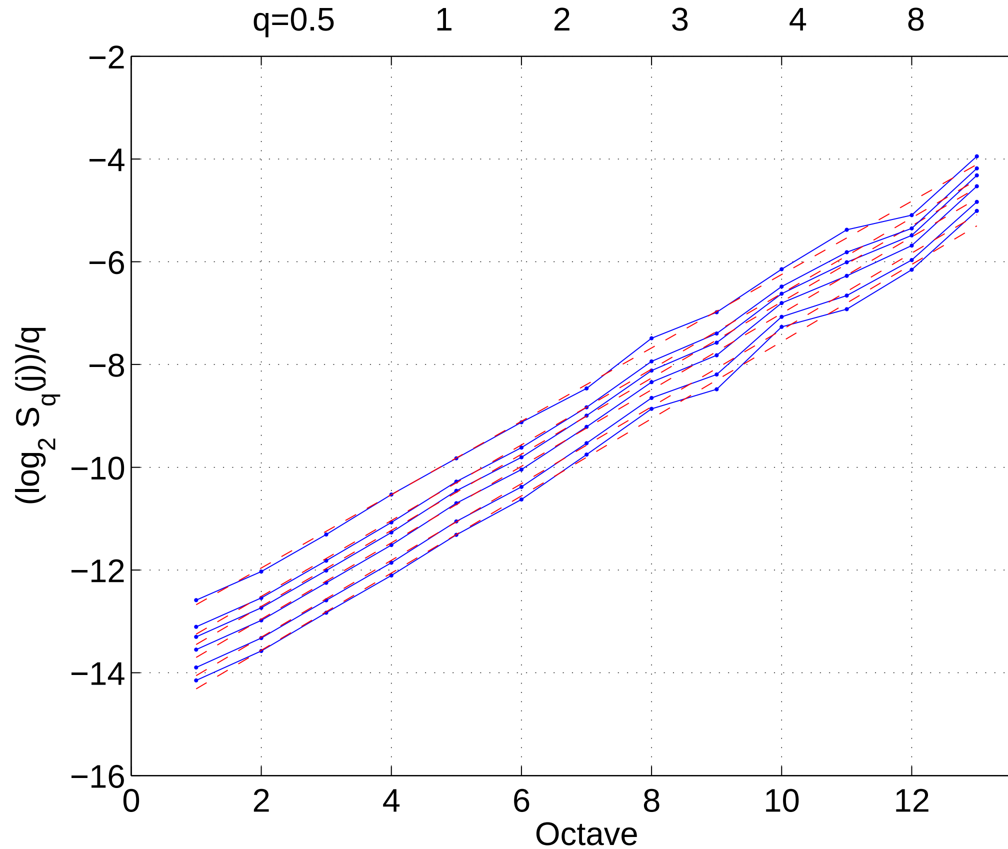
- SOLUTION : **P2 et P3** $\Rightarrow$ STATISTICAL AVERAGES $\Rightarrow$ TIME AVERAGES,
$S_q(j) = (1/n_j) \sum_{k=1}^{n_j} |d_X(j,k)|^q$

LOG-SCALE DIAGRAMS: $\log_2 S_q(j)$ *vs* $\log_2 2^j = j$

LogScale Diagrams - MultiFractal Proc.

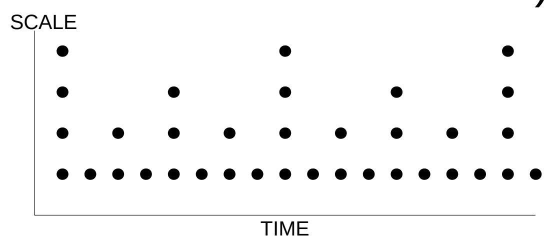






Wav. and higher-order Scaling: Estimation

- Dyadic Grid (Discrete Wavelet Transform): $a_j = 2^j, t_{j,k} = k2^j,$



- Structure Functions (Time Average):

$$S_q(j) = (1/n_j) \sum_{k=1}^{n_j} |d_X(j,k)|^q$$

- Definition:

$$Y_{j,q,n} = \log_2 S_n(2^j, q; f_0) \text{ versus } \log_2 2^j = j,$$

$$\hat{\zeta}(q,n) = \sum_{j=j_1}^{j_2} w_{j,q} Y_{j,q,n} \ .$$

Non Weigthed: $a_j = cste$

- What Are the Performance of Such Estimators ?

When applied to MultiFractal Processes

TEST FOR THE FINITENESS OF MOMENTS

GONÇALVÈS, RIEDI

THEOREM :

LET $\mathbf{X}$ BE A RV WITH CHARACTERISTIC FUNCTION $\chi(s) := \mathbb{E}\exp\{is\mathbf{X}\}$.
IF $\mathcal{H}_{\Re\chi} := \sup\{\alpha > 0 : |\Re\chi(s) - P_\alpha(s)| \leq C|s|^\alpha\}$,
IS THE LOCAL HÖLDER REGULARITY OF $\Re\chi$ AT THE ORIGIN, THEN
$\mathbb{E}|\mathbf{X}|^q < +\infty \ \forall q \leq q_c^+$ AND $\mathcal{H}_{\Re\chi} \leq q_c^+ \leq \lfloor \mathcal{H}_{\Re\chi} \rfloor + 1$.

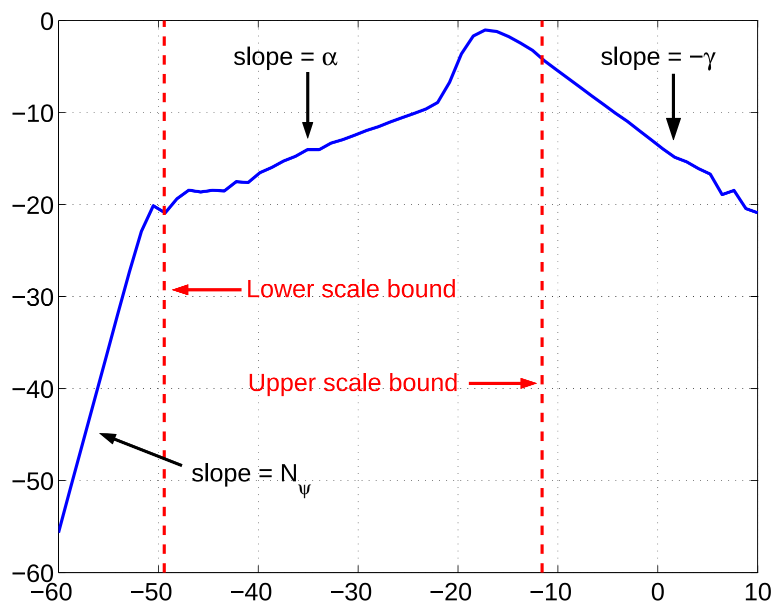
ESTIMATOR :

$\{X_k\}_{k=1,\dots,n}$, n I.I.D RVS, SET

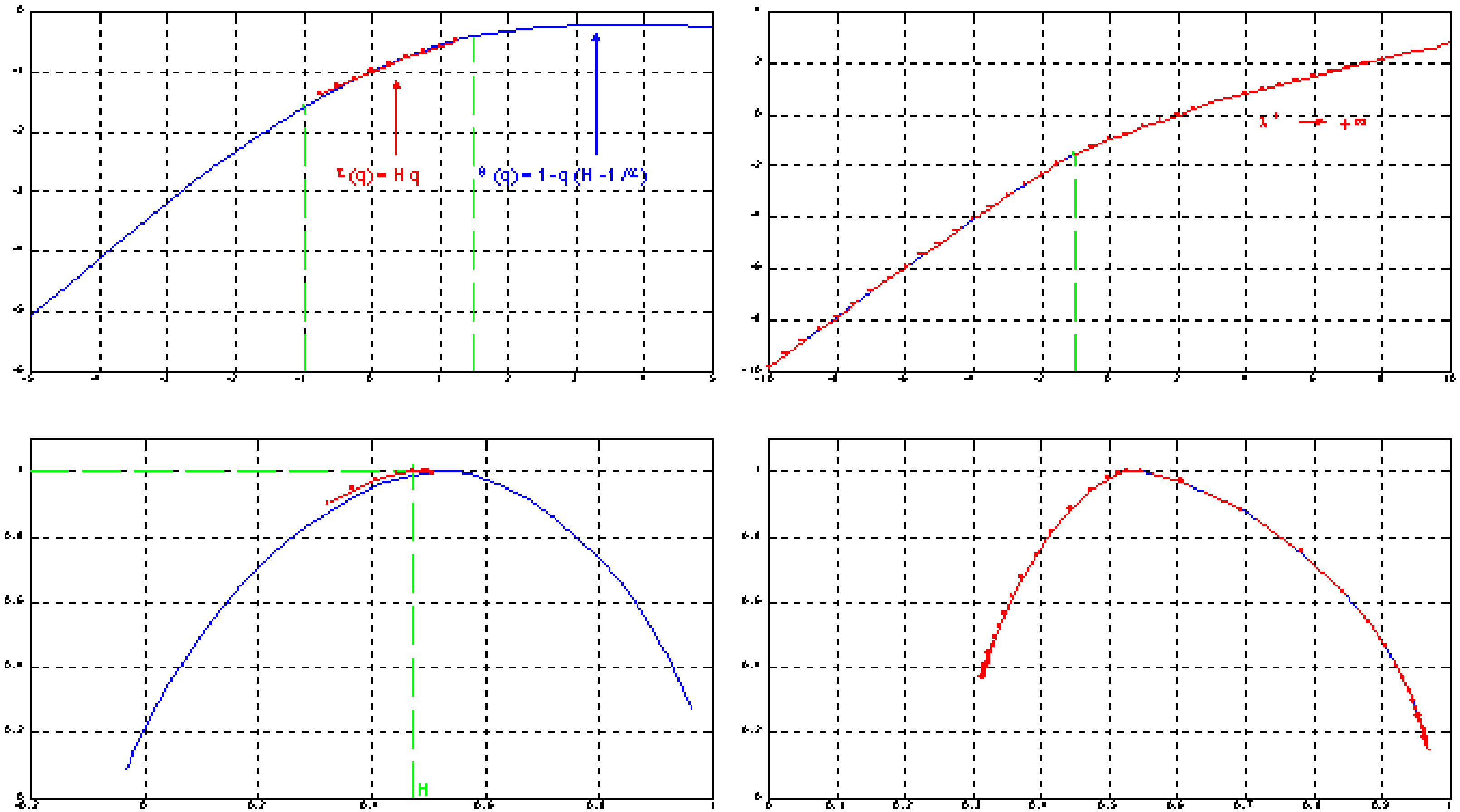
$$W(a) := n^{-1} \sum_{k=1}^{n} \Psi(a.X_k),$$

WITH Ψ A REAL AND SEMI-DEFINITE FOURIER TRANSFORM OF A SUFFICIENTLY REGULAR WAVELET ψ.
THEN

$$\mathcal{H}_{\Re\chi} = \limsup_{a \to 0^+} \frac{\log|W(a)|}{\log a}.$$



ESTIMATING THE PARTITION FUNCTION SUPPORT



METHODOLOGY

- NUMERICAL SYNTHESIS OF PROCESSES:

– ACCUMULATE $nbreal$ NUMERICAL REPLICATIONS WITH LENGTH n SAMPLES.

- APPLY SCALING EXPONENTS ESTIMATORS:

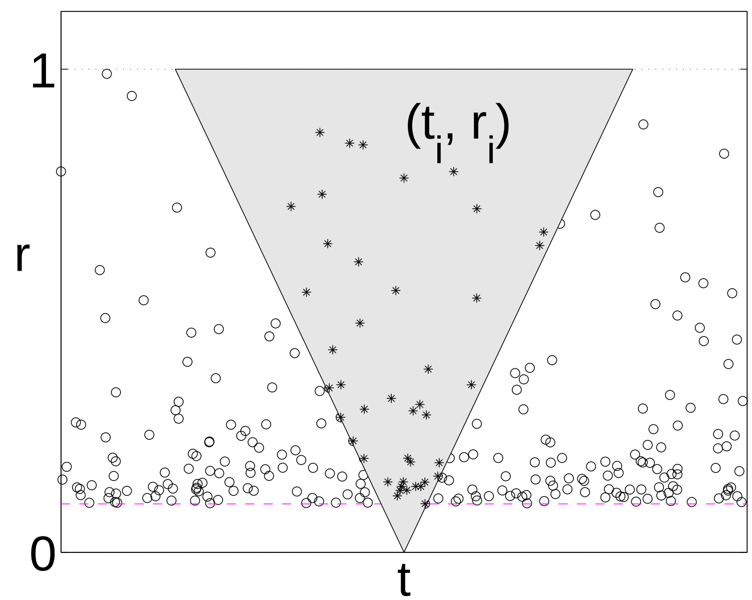
– COMPUTE $\hat{\zeta}(q,n)_{(l)}$ FOR EACH REPLICATION,

– AVERAGE OVER REPL. TO OBTAIN THE STATISTICAL PERFORMANCE OF $\hat{\zeta}(q,n)$

- ASYMPTOTIC BEHAVIOURS:

– THE CASCADE DEPTH INCREASES FOR A GIVEN NUMBER OF INTEGRAL SCALES.

– ... ,



METHODOLOGY

- NUMERICAL SYNTHESIS OF PROCESSES:

— ACCUMULATE $nbreal$ NUMERICAL REPLICATIONS WITH LENGTH n SAMPLES.

- APPLY SCALING EXPONENTS ESTIMATORS:

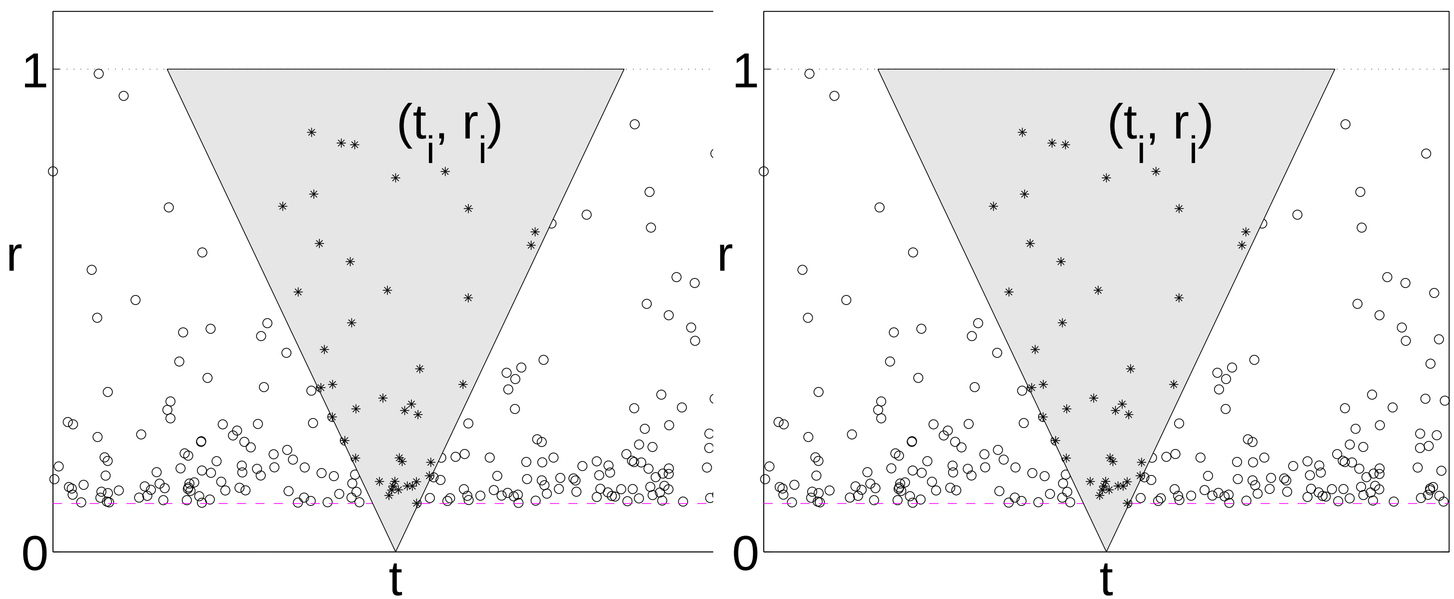
— COMPUTE $\hat{\zeta}(q,n)_{(l)}$ FOR EACH REPLICATION,

— AVERAGE OVER REPL. TO OBTAIN THE STATISTICAL PERFORMANCE OF $\hat{\zeta}(q,n)$

- ASYMPTOTIC BEHAVIOURS:

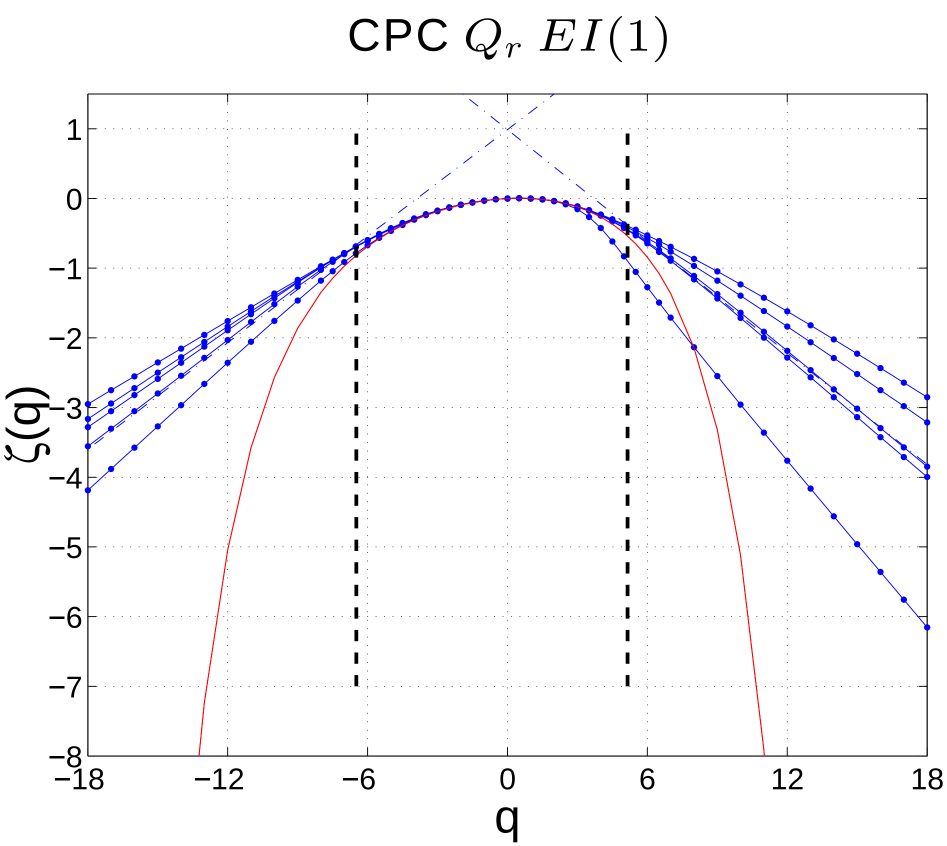
— THE CASCADE DEPTH INCREASES FOR A GIVEN NUMBER OF INTEGRAL SCALES.

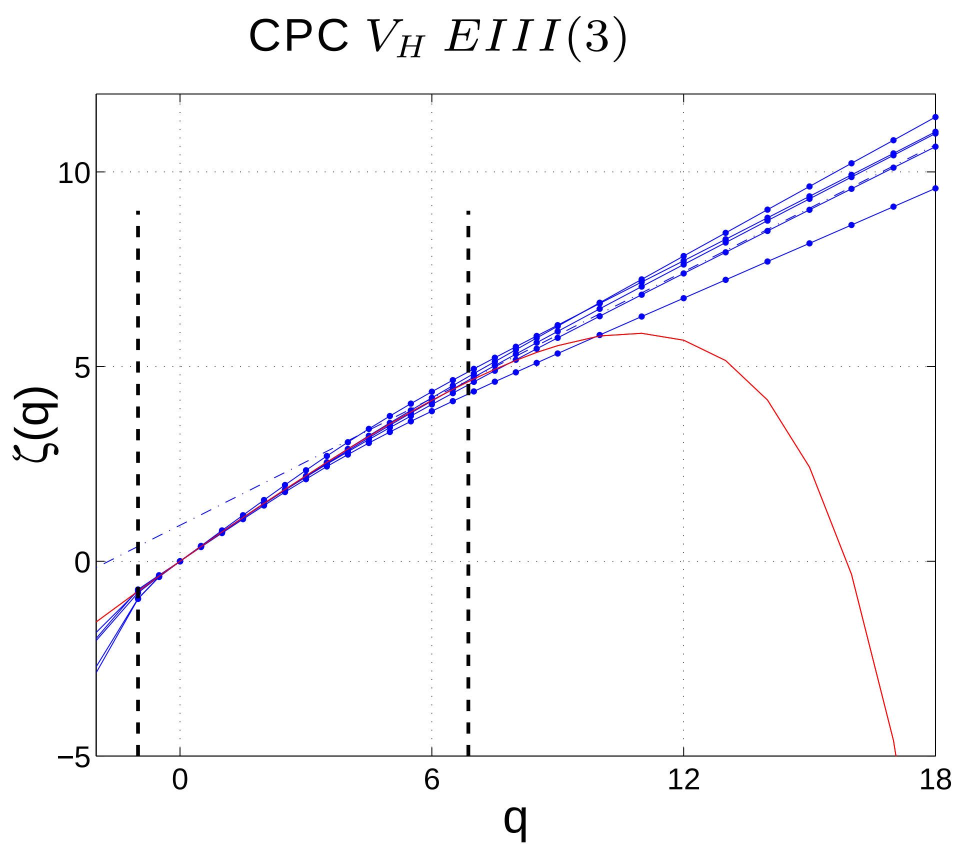
— THE NUMBER OF INTEGRAL SCALES INCREASES FOR A GIVEN CASCADE DEPTH,



LINEARISATION EFFECT: $\hat{\zeta}(q)$

LASHERMES, ABRY, CHAINAIS



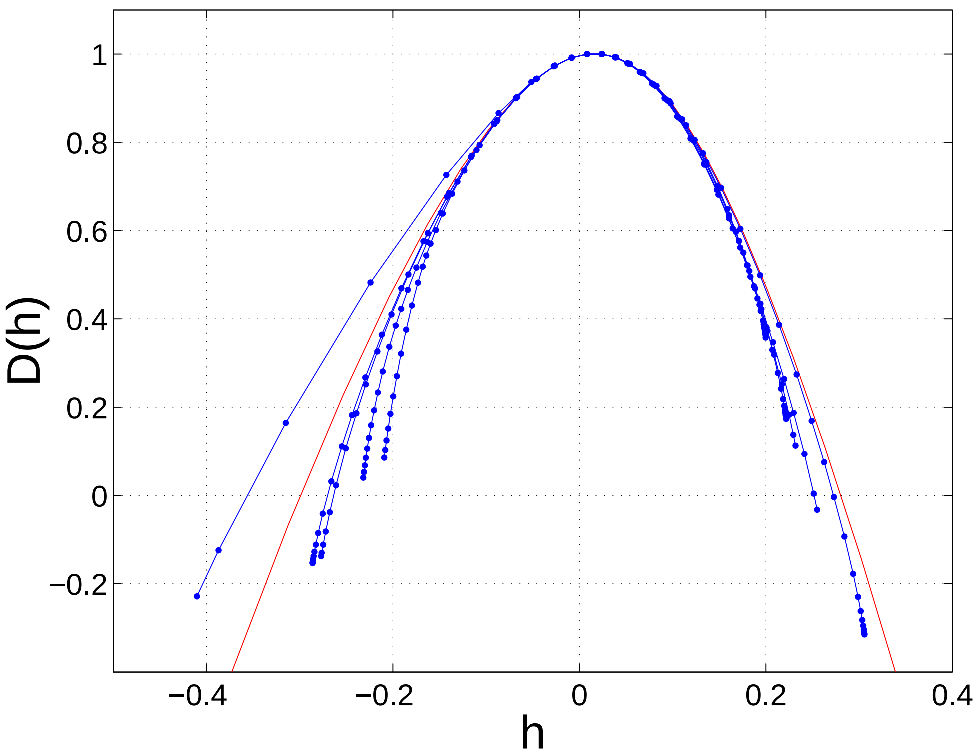


$q > q_o$, $\hat{\zeta}(q,n) = \alpha_o + \beta_o q$, q_o, α_o, β_o ARE RV.

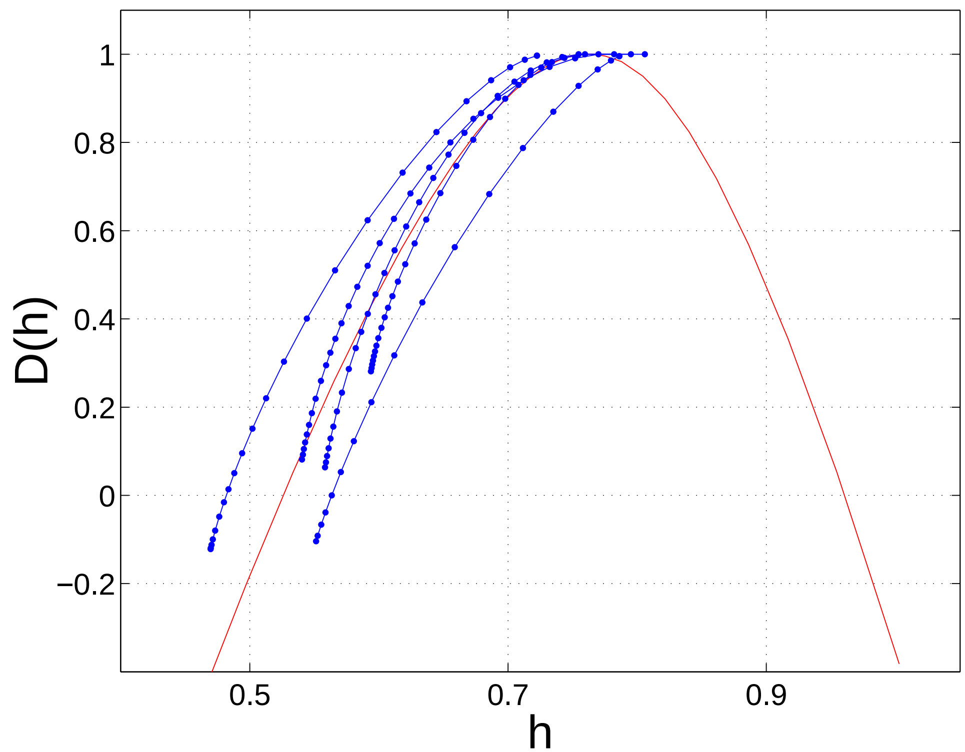
LINEARISATION EFFECT: LEGENDRE TRANSFORM

$D(h) = d + \text{MIN}_q(qh - \zeta(q))$, ($d$ EUCLIDIEN DIMENSION OF SPACE).

CPC Q_r $EI(1)$



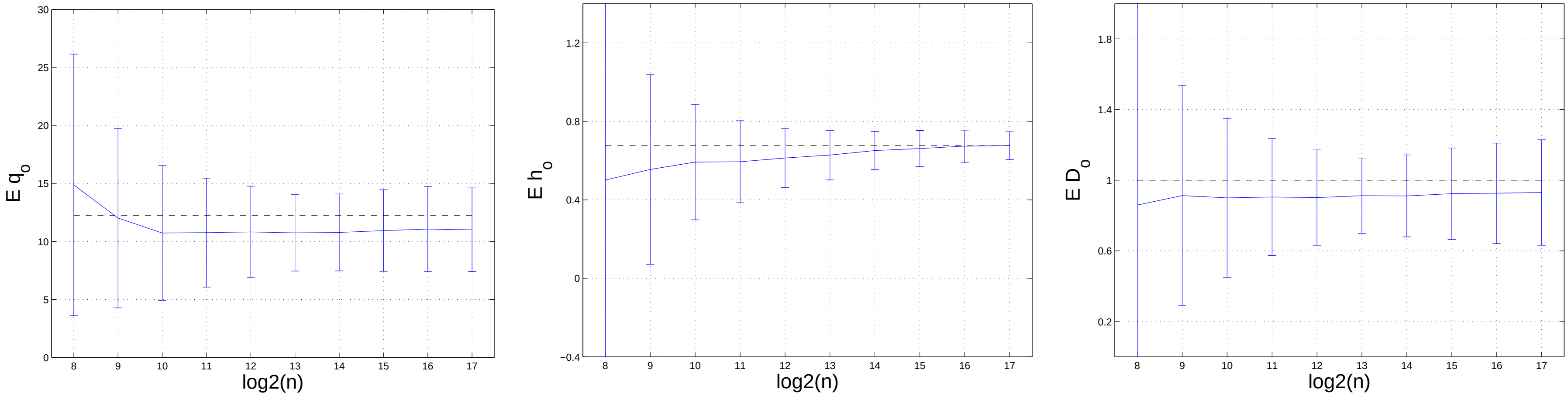
CPC V_H $EIII(3)$



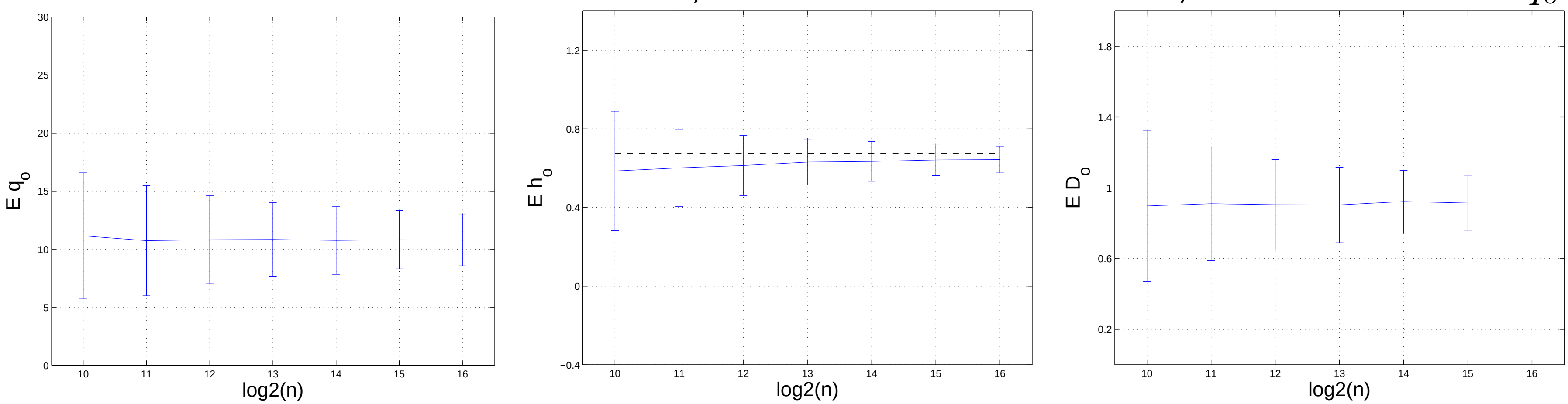
ACCUMULATION POINTS : $D_o(h_o)$, WITH $D_o = d - \alpha_o,\ \ h_o = \beta_o$,
D_o, h_o ARE RV.

Lin. Effect: Asymptotic behaviours

- Given resolution, Increasing Number of Integral Scales, q_0



- Given Number of Integral Scales, Increasing Resolution, q_0



LINEARISATION EFFECT: CONJECTURE

- CRITICAL POINTS:

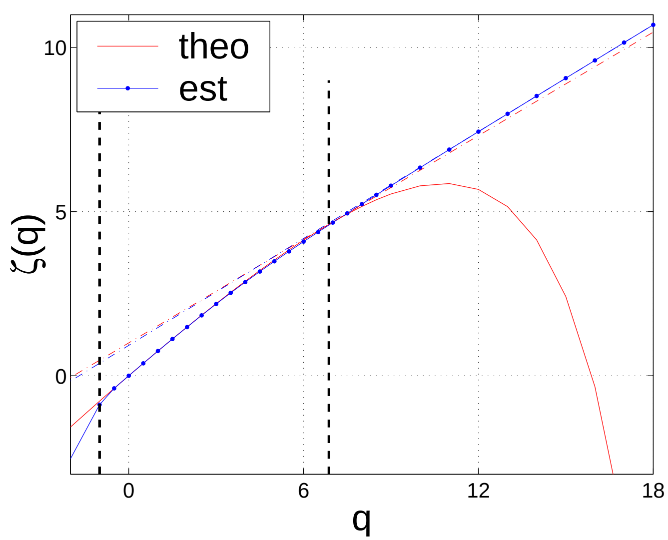
$$\begin{cases} D_*^{\pm} & = & 0, \\ D(h_*^{\pm}) & = & 0, \\ h_*^{\pm} & = & (d\zeta(q)/dq)_{q=q_*^{\pm}}. \end{cases}$$

- RESULTS:

$$EI : \begin{cases} \hat{\zeta}(q,n) = d - D_o^- + h_o^- q & \longrightarrow & d - D_*^- + h_*^- q, & q \le q_*^-, \\ \hat{\zeta}(q,n) & \longrightarrow & \zeta(q), & q_*^- \le q \le q_*^+, \\ \hat{\zeta}(q,n) = d - D_o^+ + h_o^+ q & \longrightarrow & d - D_*^+ + h_*^+ q, & q_*^+ \le q. \end{cases}$$

$$EII\&III : \begin{cases} \hat{\zeta}(q,n) & \longrightarrow & \zeta(q), & 0 < q \le q_*^+, \\ \hat{\zeta}(q,n) = d - D_o^+ + h_o^+ q & \longrightarrow & d - D_*^+ + h_*^+ q, & q_*^+ \le q. \end{cases}$$

- ILLUSTRATION:

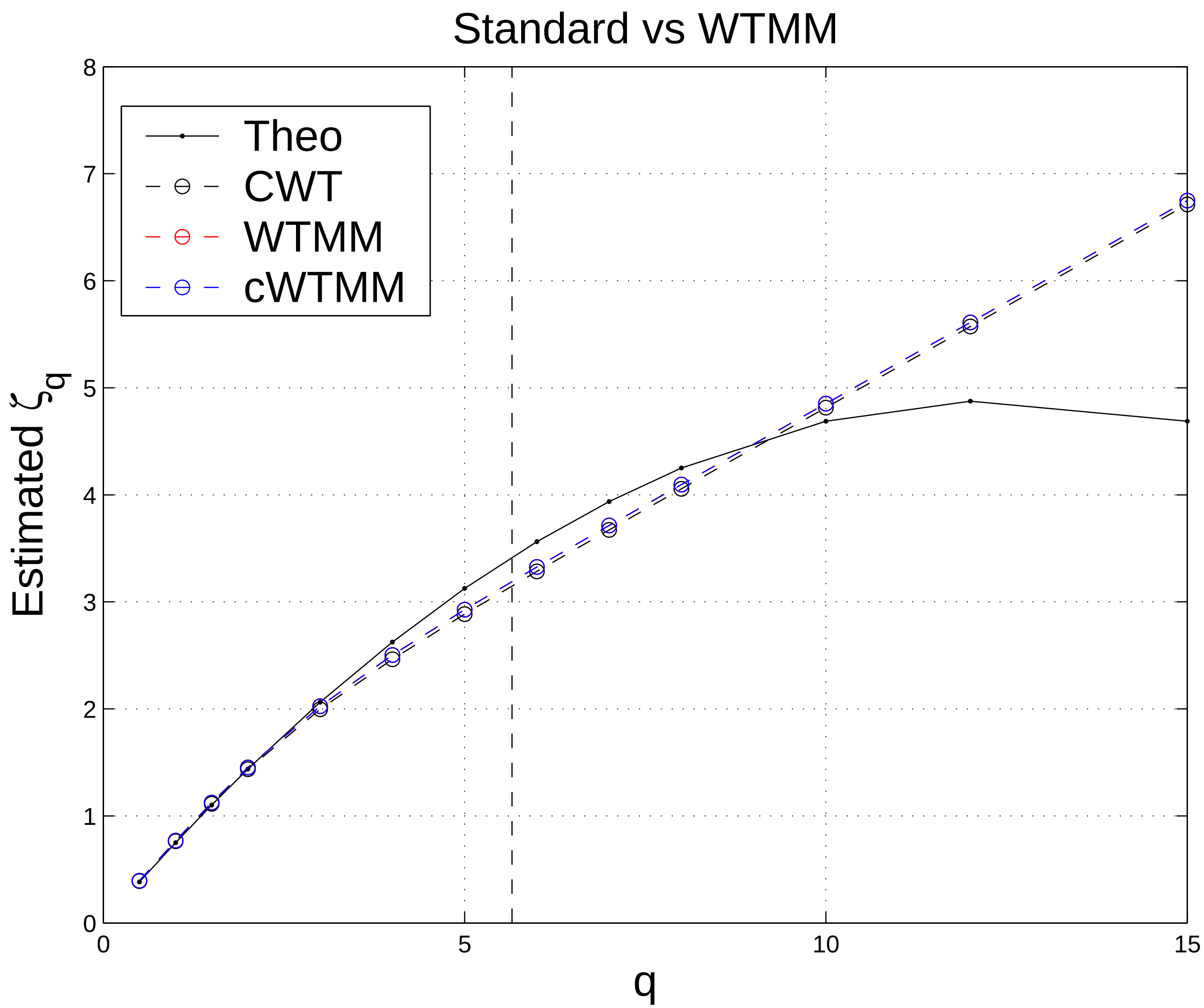


LINEARISATION EFFECT: COMMENTS

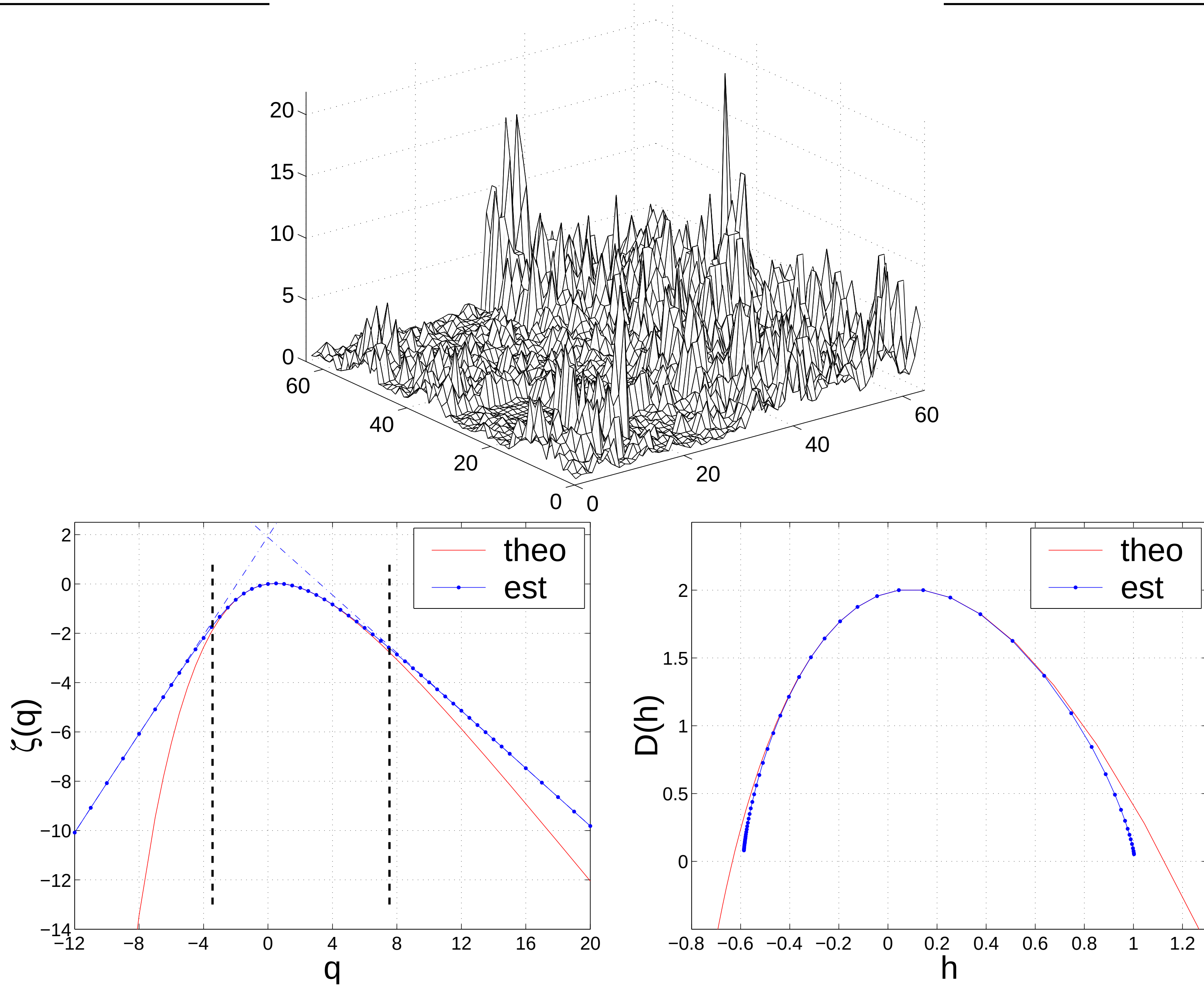
WHEN DOES THE LINEARISATION EFFECT EXIST ?

- FOR ALL TYPES OF CASCADES: CMC, CPC, IDC,
- FOR ALL TYPES OF PROCESSES: Q_r, A, V_H, Y_H,
- FOR ALL NUMBERS OF VANISHING MOMENTS: $N \geq 1$,
- FOR ALL MRA-BASED ESTIMATORS: WAVELETS, INCREMENTS, AGGREGATION,
- CAN BE WORKED OUT FOR $q < 0$,
- EXTENDS TO DIMENSION HIGHER THAN $d > 1$.

Extension: Standard WT versus WTMM (1/3).

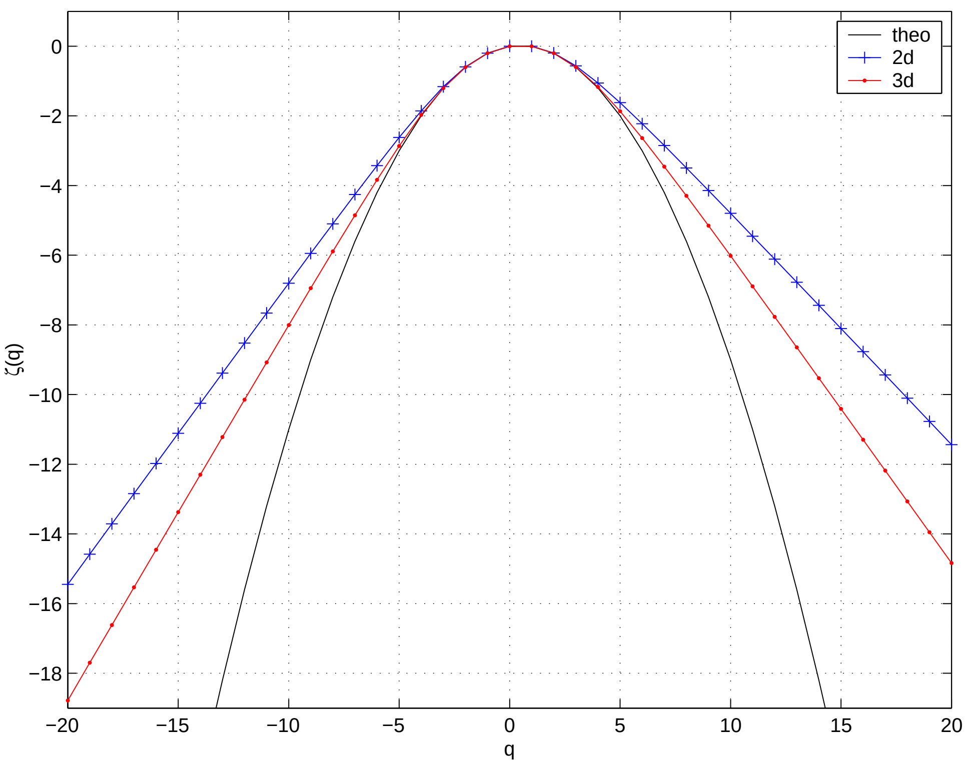


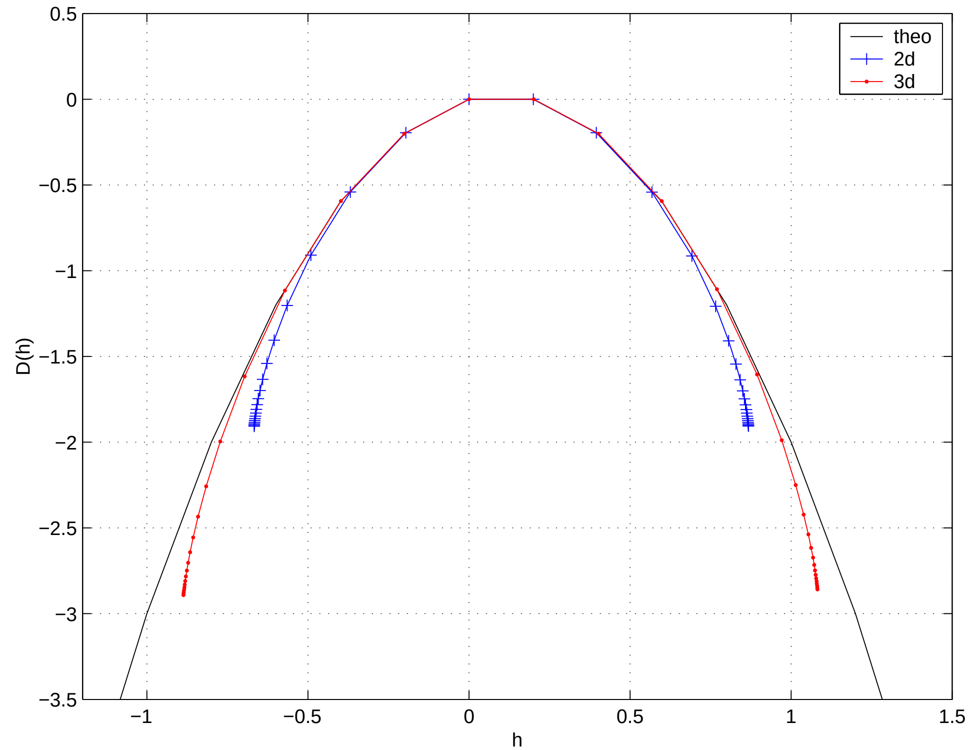
EXTENSION: 2D MULTIPLICATIVE CASCADE (2/3).



EXTENSION: 3D MULTIPLICATIVE CASCADE (3/3).

3D CMC (LOG NORMAL), EI(1) COMPARED TO A 2D SLICE.





Linearisation Effect: Comments

When does the Linearisation Effect EXIST ?

— for all types of cascades: CMC, CPC, IDC,

— for all types of processes: Q_r, A, V_H, Y_H,

— for all numbers of Vanishing Moments: $N \geq 1$,

— for all MRA-based estimators: Wavelets, Increments, Aggregation,

— can be worked out for $q < 0$,

— extends to dimension higher than $d > 1$.

What the Linearisation Effect IS NOT:

— a low performance estimation effect.

— a finite size effect : The critical parameters do not depend on n,

be it the number of integral scales,

or the depth (or resolution) of the cascades.

— a finiteness of moments effect,

- $q_c^- < 0 < 1 < q_c^+, \;\; q - 1 + \varphi(q) = 0$,
- $q_c^- < q_*^- < 0 < 1 < q_*^+ < q_c^+$,

What the Linearisation Effect Might Be:

— Multiplicative Martingales ?

— Ossiander, Waymire 00, Kahane, Peyrière 75, Barral, Mandelbrot 02.

Linearisation Effect: Picture

- Two power-laws, two functions of q:

– Bare Cascade:
$$\mathbb{E}Q_r(t)^q = r^{\varphi(q)}, \quad q \in \mathcal{R}.$$

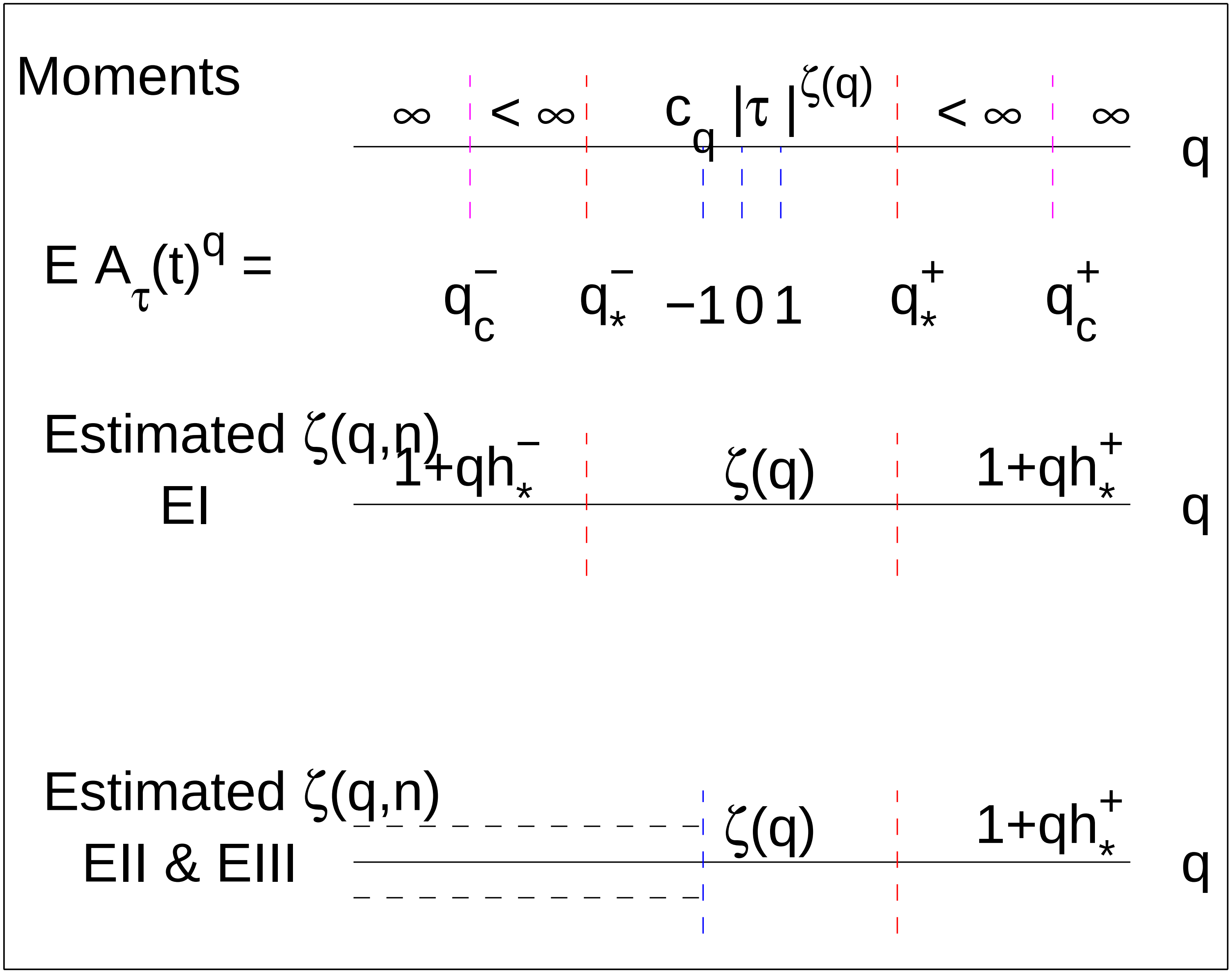
– Dressed cascade:
$$\left.\begin{array}{rcll} \mathbb{E}T_{Q_0}(t,a;\beta_0)^q & = & c_q|a|^{\zeta(q)}, & q \in [q_c^-, q_c^+], \\ \mathbb{E}T_{Q_0}(t,a;\beta_0)^q & = & \infty, & \text{else}, \end{array}\right\}$$

with:
$$\left.\begin{array}{rcll} \zeta(q) & = & 1 + qh_*^-, & q \in [q_c^-, q_*^-], \\ \zeta(q) & = & \varphi(q), & q \in [q_*^-, q_*^+], \\ \zeta(q) & = & 1 + qh_*^+, & q \in [q_*^+, q_c^+]. \end{array}\right\}$$

- Confusion between $\varphi(q)$ and $\zeta(q)$:

– Multiplicative Cascade: $\varphi(q), q \in \mathcal{R}$,
– Scaling Exponents: $\zeta(q), q \in [q_c^-, q_c^+]$.

LINEARISATION EFFECT: SKETCHED VIEWS



Linearisation Effect: Impacts and Importance

Consequences: Recast the Usual Goals :

– Estimate the Integral Scale and the Resolution of the Cascade,
$\Rightarrow$ i.e., Find a Scaling Range $[a_m, a_M]$
– Estimate the Critical Parameters $D_*^\pm, h_*^\pm, q_*^\pm$,
– Estimate the $\zeta(q)$ for $q \in [q_*^-, q_*^+]$,
$\longrightarrow$ Visit B. Lashermes's Poster.

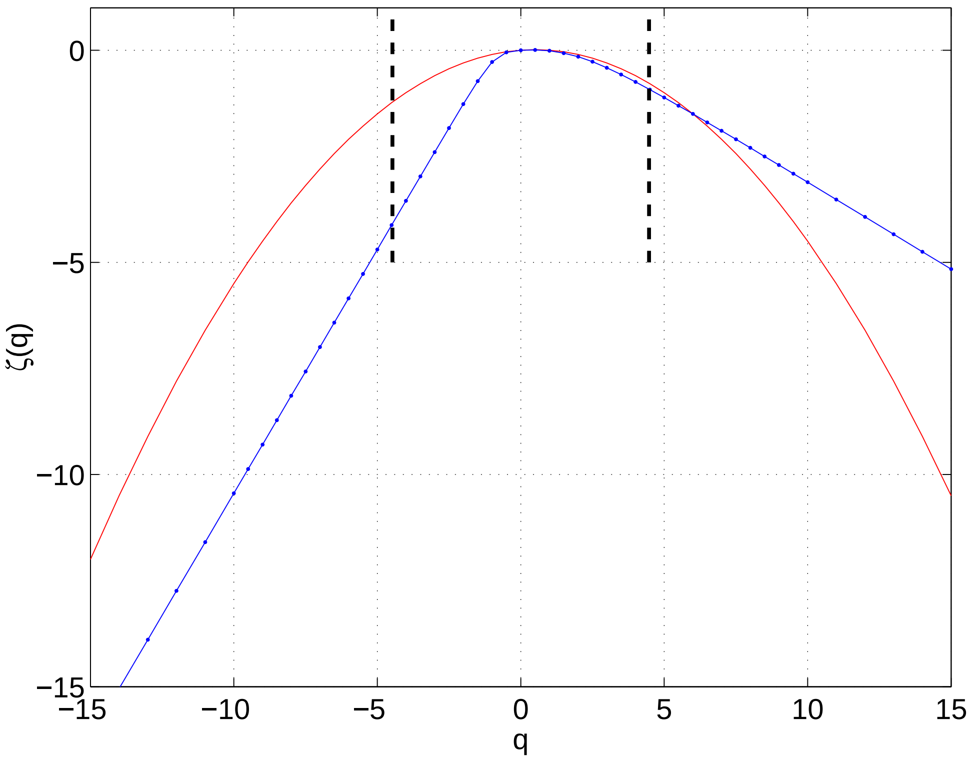
Importance of the Linearisation Effect:

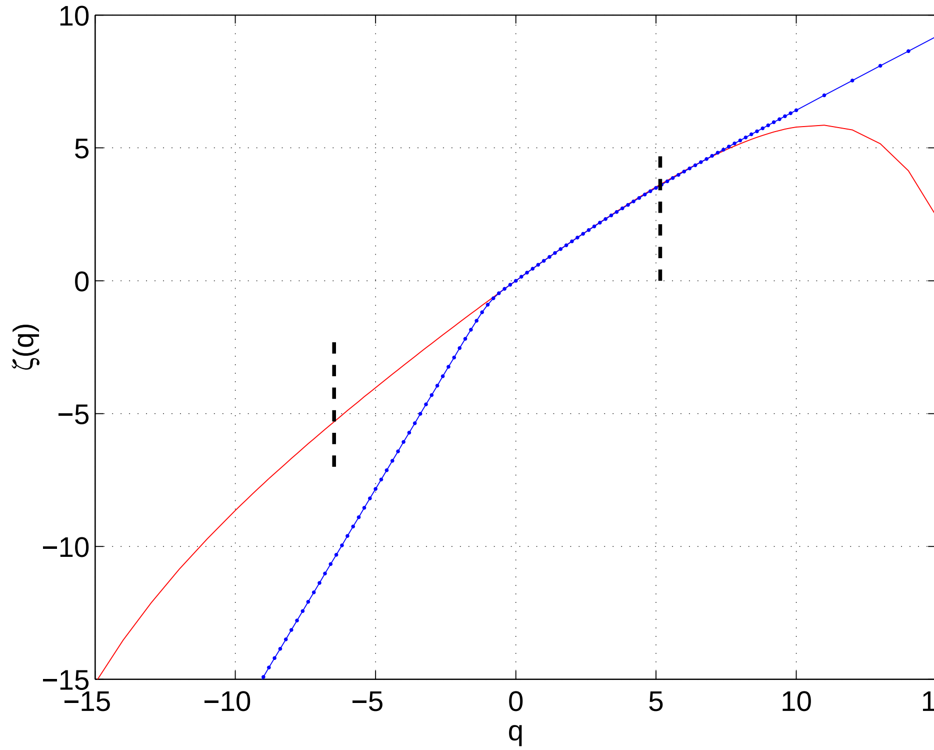
– Discrimination of MF Models based on $\hat{\zeta}(q, n)$,
– Discrimination between monoFractal and MultiFractal,

NEGATIVE VALUES OF qS

DIFFICULTIES ?

- FINITENESS ? $S_q(j) = (1/n_j)\sum_{k=1}^{n_j} |d_X(j,k)|^q < \infty$?
- NUMERICAL INSTABILITY ? $d_X(j,k) \simeq 0 \rightarrow |d_X(j,k)|^q = \infty$
- THEORY ? FULL MULTIFRACTAL SPECTRUM

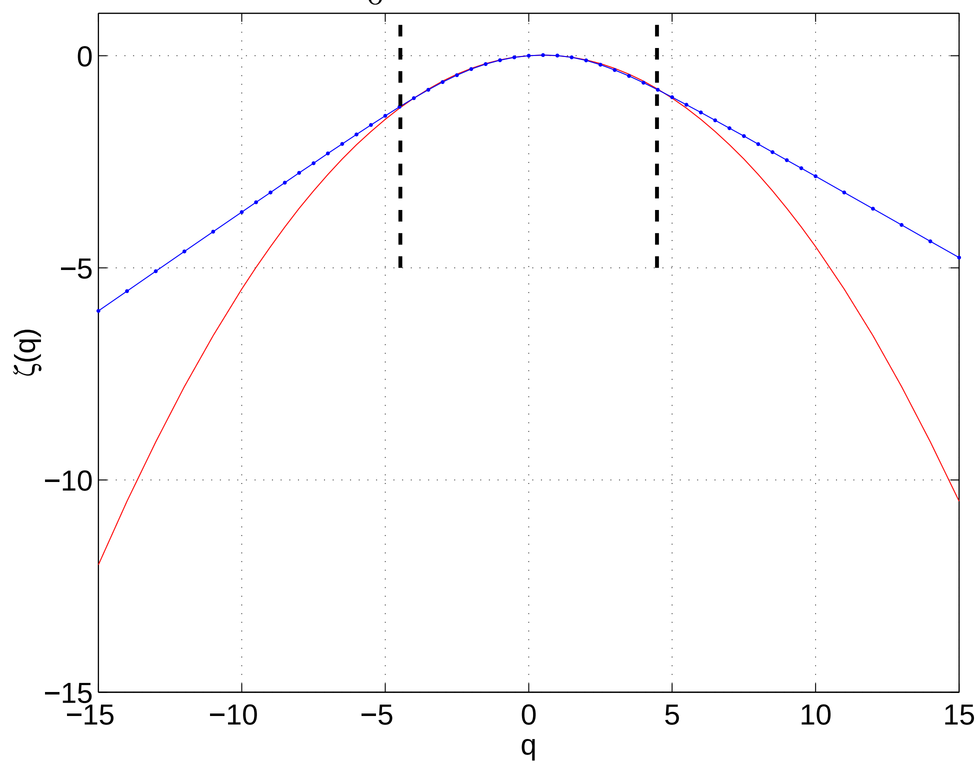




SOLUTIONS ?

NEGATIVE VALUES OF qS - SOLUTION 1

AGGREGATION: $T_X(a,t) = \frac{1}{aT_0}\int_t^{t+aT_0} X(u)du$

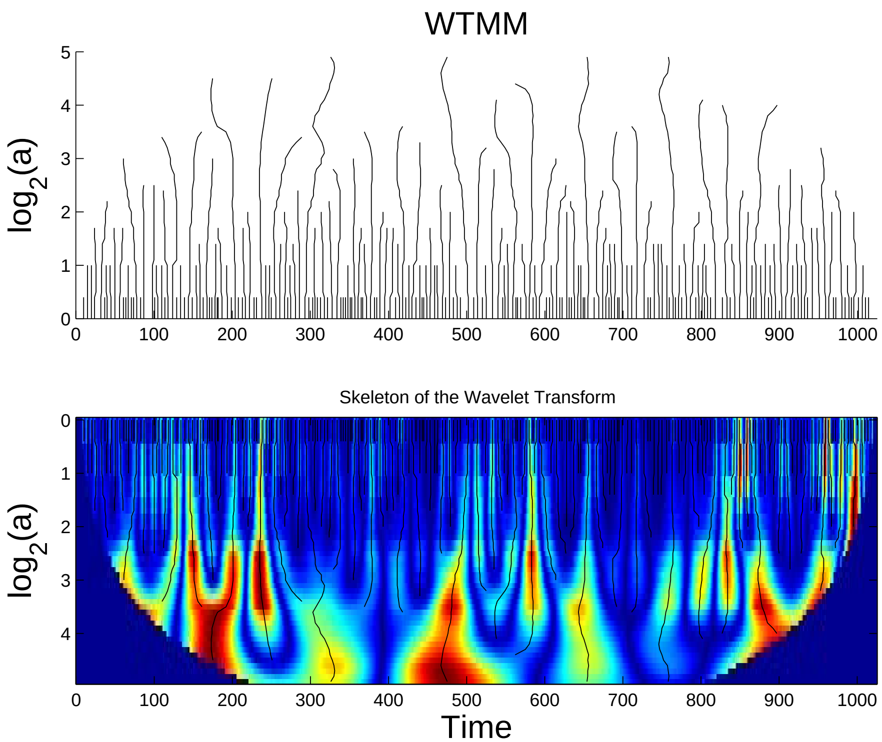


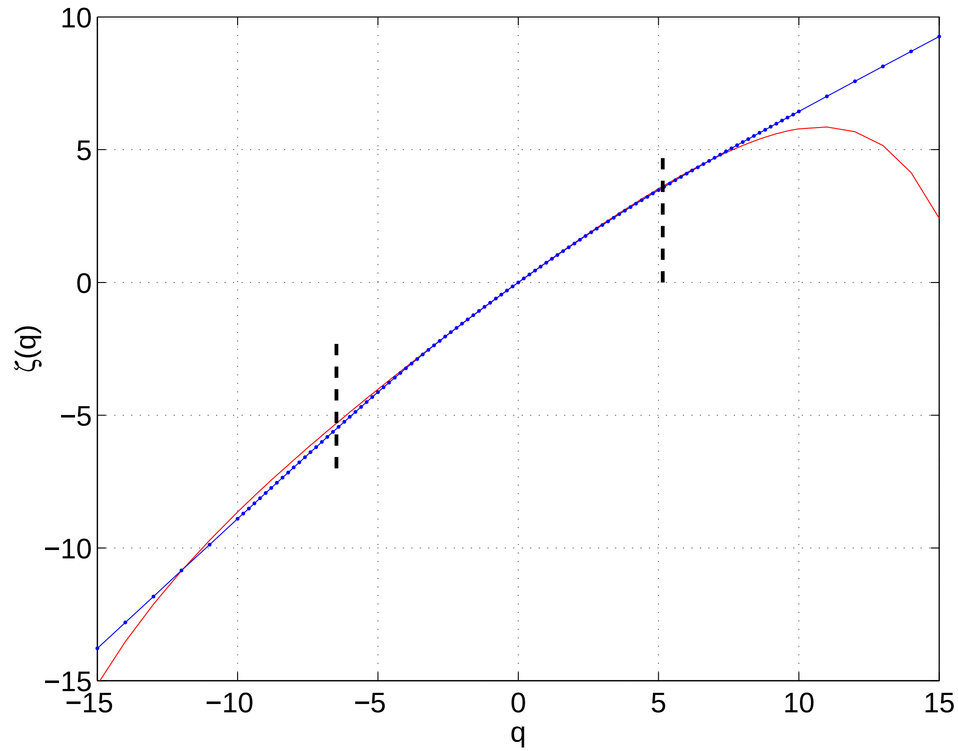
APPLIES ONLY TO POSITIVE DATA (MEASURE)

Negative Values of qs - Solution 2

WT Modulus Maxima (Arneodo et al.)

$$L_X(a, t_k) = \text{SUP}_{a'<a}|T_X(a', t_k(a'))|$$



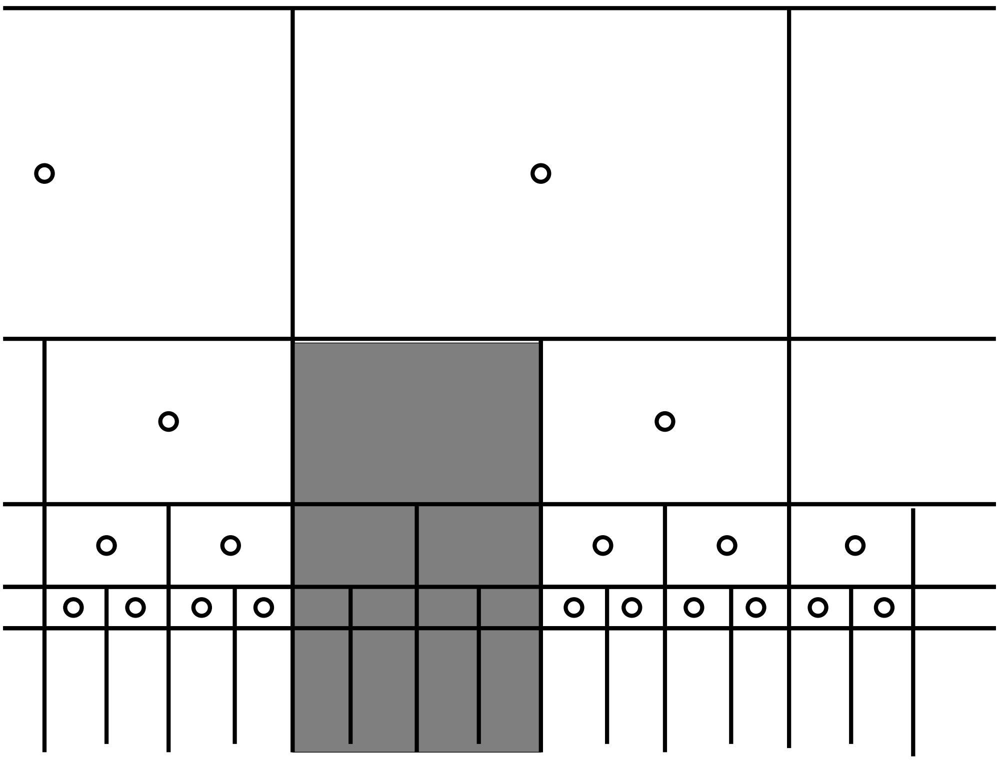


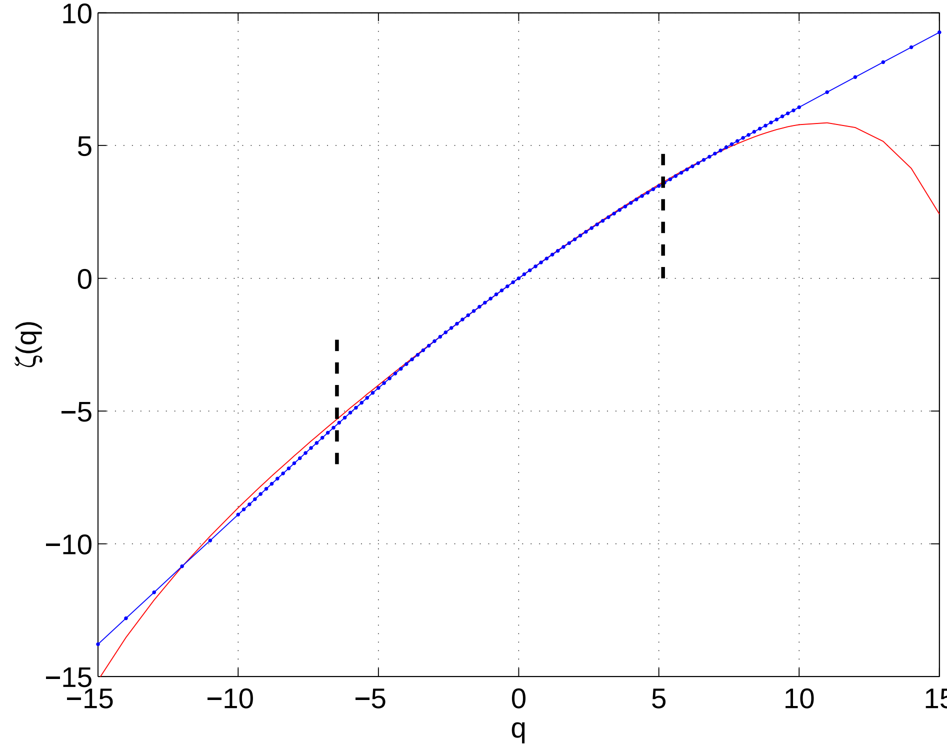
Computationally Expensive

NEGATIVE VALUES OF qS - SOLUTION 3

WAVELET LEADERS: (JAFFARD ET AL.)

$$d_X(j,k) \rightarrow L_X(j,k) = \text{SUP}_{j'<j} d_X(j', 2^{-j'})$$





COMPUTATIONALLY EFFICIENT AND EXCELLENT STATISTICAL PERFORMANCE

BEYOND POWER LAWS

- SELF-SIMILARITY:

$$\mathbb{E}|d_X(j,k)|^q = C_q (2^j)^{qH} = C_q \exp(qH \ln 2^j)$$

  - POWER LAWS,
  - $\forall 2^j$ (FOR ALL SCALES),
  - $\forall q / \mathbb{E}|d_X(j,k)|^q < \infty$,
  - A SINGLE PARAMETER H
  - ADDITIVE STRUCTURE.

- MULTIFRACTAL

$$\mathbb{E}|d_X(j,k)|^q = C_q (2^j)^{\zeta(q)} = C_q \exp(\zeta(q) \ln 2^j)$$

  - POWER LAWS,
  - $\forall 2^j < L$, (FOR FINE SCALES ONLY, IN THE LIMIT $2^j \to 0$,)
  - $\forall q$?
  - A WHOLE COLLECTION OF SCALING PARAMETER $\zeta(q)$
  - MULTIPLICATIVE STRUCTURE.

- BEYOND POWER LAWS : WARPED INF. DIV. CASCADES

$$\mathbb{E}|d_X(j,k)|^q = C_q (2^j)^{qH} = C_q \exp(qH \ln 2^j)$$
$$\mathbb{E}|d_X(j,k)|^q = C_q (2^j)^{\zeta(q)} = C_q \exp(\zeta(q) \ln 2^j)$$
$$\mathbb{E}|d_X(j,k)|^q = \quad = C_q \exp(\zeta(q) n(2^j))$$

$\longrightarrow$ VISIT PIERRE CHAINAIS'S POSTER

CONCLUSIONS AND REFERENCES

ANALYSING SCALING IN DATA ?

– THINK WAVELET

— EFFICIENCY,

— PRACTICAL AND CONCEPTUAL ADEQUATION AND SIMPLICITY,

— ROBUSTNESS AGAINST NON STATIONARITIES,

— EASY TO USE, LOW COAST, REAL TIME ON LINE.

MODELLING SCALING IN DATA ?

– THINK SELF SIMILARITY VERSUS MULTIPLICATIVE CASCADES,

– AND POSSIBLY ADD LONG MEMORY.

– ALSO SCALING MAY NOT BE POWER LAWS

REFERENCES AND RESOURCES, VISIT :

— **perso.ens-lyon.fr/patrice.abry**

— **inrialpes.fr/is2/∼pgoncalv**

— **www.cubinlab.ee.mu.oz.au/∼darryl**

— **fraclab**

— **www.isima.fr/∼chainais**